Analyse numérique d'équations aux dérivées aléatoires, applications à l'hydrogéologie

Julia Charrier

THÈSE / ENS CACHAN - BRETAGNE

sous le sceau de l'Université européenne de Bretagne

pour obtenir le titre de

DOCTEUR DE L'ÉCOLE NORMALE SUPÉRIEURE DE CACHAN

Mention : Mathématiques

École doctorale MATISSE

présentée par

Julia Charrier

Préparée à l'INRIA Rennes - Bretagne Atlantique

Analyse numérique d'équations aux dérivées partielles à coefficients aléatoires, applications à l'hydrogéologie.

Thèse soutenue le 12 juillet 2011

devant le jury composé de :

Fabio Nobile
Professeur à Politecnico di Milano / *rapporteur*

Denis Talay
Directeur de recherche à l'INRIA / *rapporteur*

Olivier Le Maître
Chargé de recherche CNRS au LIMSI / *examinateur*

Florent Malrieu
Maître de conférences à l'université de Rennes 1 / *examinateur*

Arnaud Debussche
Professeur à l'ENS Cachan-Bretagne / *directeur de thèse*

Jocelyne Erhel
Directrice de recherche à l'INRIA / *directrice de thèse*

Remerciements

Mes premiers remerciements vont tout naturellement à mes directeurs de thèse Arnaud et Jocelyne pour avoir accepté d'encadrer ma thèse. Tout au long de ces trois ans, ils ont su me guider et me conseiller tout en ma laissant une grande autonomie et en me faisant confiance. Je remercie tout particulièrement Arnaud pour sa grande disponibilité, son soutien sans faille, toutes les connaissances qu'il m'a transmises, pour m'avoir fait profiter de son expérience, pour sa bienveillance dans les moments difficiles, pour avoir cru en moi et pour m'avoir fait avancer sur le chemin qui mène d'étudiante à chercheuse.

Je remercie Denis Talay et Fabio Nobile d'avoir accepté de rapporter cette thèse, malgré leurs emplois du temps très chargés. C'est un honneur pour moi qu'ils aient lu avec attention mon travail et qu'ils donnent ainsi un point de vue spécialisé sur celui-ci.

Je remercie Florient Malrieu et Oliver Le Maître d'avoir accepté de faire partie de mon jury. Je leur suis très reconnaissante pour l'intérêt qu'ils portent à mon travail.

Je remercie Robert Scheichl, Ivan Graham et Aretha Teckentrup pour leur accueil très chaleureux lors de mon séjour à Bath, et avec qui j'ai travaillé avec plaisir et beaucoup appris.

Je tiens également à remercier tous les membres du département de maths de Ker Lann, pour leurs qualités humaines, pédagogiques et scientifiques. J'ai vraiment beaucoup apprécié l'ambiance de travail chaleureuse qui y règne. Je tiens en particulier à exprimer toute ma gratitude à Michel Pierre, qui a été un excellent professeur pour moi, et à Grégory Vial, dont la gentillesse et les qualités pédagogiques sont remarquables, pour la disponibilité et l'efficacité dont ils ont tous deux fait preuve à chaque fois que je suis allée les voir avec des questions au cours de ma thèse. Merci à Erwan Faou et Philippe Chartier pour leur gentillesse et leur humeur joviale ainsi qu'à Grégory Vial, j'ai vraiment eu plaisir à travailler avec eux trois dans le cadre de mes enseignements. Merci à Virginie Bonnaillie-Noël pour sa gentillesse et son soutien, à Rozenn Texier-Picard pour avoir été ma tutrice de monitorat et pour sa gentillesse, à Yannick Privat avec qui on a passé des soirées sympas et à Benoît Cadre. Merci à mes co-bureaux de Ker Lann, actuels et anciens, grâce à qui l'ambiance est vraiment vivante et agréable : Guillaume, Quentin pour sa bonne humeur et les soirées sympas passées ensemble, Charlie pour sa disponibilité face à mes questions et les soirées sympas passées ensemble, Ludovic pour ses conseils, Agnès, Shanshan et Martina pour leur gentillesse et leur bonne humeur, Thibaut, Jimmy, Fanny, Sébastien. On a à la fois bien rigolé et eu des discussions de maths très intéressantes sur la plateau.

Je remercie tous les membres de l'équipe SAGE à l'IRISA, qui m'ont accueillie chaleureusement. Tout d'abord je remercie les permanents : Jocelyne, Bernard Philippe et Edouard Canot, et plus récemment Géraldine Pichot pour leur gentillesse. Je remercie également Baptiste, Nadir et Géraldine pour m'avoir aidée efficacement quand j'en ai eu besoin. Je remercie Désiré pour avoir été un co-bureau très agréable, qui m'a fait rire et est accouru à mon secours quand j'avais des problèmes d'informatique, le tout dans une ambiance studieuse et sympathique. Je remercie Mohamad pour sa gentillesse. Et enfin je remercie collectivement ceux que j'ai moins vus et donc moins connus.

Enfin j'en profite pour remercier ceux et celles qui m'ont apporté de la joie et m'ont soutenue dans ma vie personnelle ces dernière années, en particulier dans les moments sombres. Tout d'abord merci à ma soeur adorée Lauriaïs, pour sa bonne humeur, son soutien et son sens de l'humour inimitable. Un grand merci à

Elodie, qui est ma meilleure amie depuis maintenant plus de vingt ans, et qui, même si on est très différentes, est là pour m'écouter, me soutenir, sans porter de jugement depuis tout ce temps. J'espère de tout coeur qu'on résussira à faire durer et mûrir cette belle amitié. Merci également à Maud pour son amitié, son soutien, sa gentillesse et sa joie de vivre. J'en profite pour remercier également Raphaël pour sa gentillesse, et à tous les deux pour m'accueillir à Paris régulièrement ! Merci à Marie pour son amitié, la vie nous a un peu éloignées, pour avoir été là plusieurs fois quand j'avais cruellement besoin d'aide, pour m'avoir redonné le goût de la cuisine, et pour les vacances passées ensemble, que j'ai vraiment appréciées. Merci à Géraldine pour son amitié, pour les randos en Bretagne faites et à faire dont je garde un super souvenir, pour son soutien et sa serviabilité. Merci à Gilles pour toutes ces super discussions pendant nos poses thé/clopes à l'IRISA et son soutien efficace quand je doutais beaucoup pour ma thèse. Merci à Jimmy pour avoir répondu à mes questions de maths, m'avoir soutenue au début de ma thèse, et encore maintenant. Merci à Guillaume et Thibaut pour avoir rendu le plateau plus vivant, mais surtout pour avoir été mes amis. Vous avez été beaucoup présents dans les moments les plus difficiles et en particulier je ne sais pas comment j'aurais fini ma thèse sans vous. Merci aussi pour toutes les soirées bien sympas qu'on a faites ensemble dans les moments moins sombres. Merci à Mouton pour les soirées sympas qu'on a passées ensemble et pour sa gentillesse. Merci à Cyrille pour toutes les discussions sympas et très intéressantes qu'on a eu ces derniers temps, ainsi que pour sa bienveillance. Merci à Ludo et Sarah pour les (trop peu nombreux) bons moments partagés avec eux cette année, pour leur soutien et leur bienveillance, en espérant en partager d'autres à l'avenir. Plus ponctuellement, merci à Antoine pour sa compréhension et son soutien, merci à PJ pour sa grâce et pour m'avoir donné envie de me mettre à la guitare, merci à Manu pour la cure de rire et merci à François pour sa douceur, sa finesse et son soutien.

Pour finir, merci à tous les auteurs, chanteurs et réalisateurs, qui ont laissés des trésors : livres, films, chansons qui ont été pour moi des étoiles, me permettant de ne pas me perdre, et m'apportant des moments de lumière et de grâce dans les nuits les plus noires. Je pense par exemple à Clarissa Pinkola Estes, Alice Miller, Krysztof Kieslovski, Hayao Miyazaki, Lars Von Trier, Mano Solo...

Table des matières

Chapitre 1

Introduction

1.1 Préliminaires : quantification des incertitudes

1.1.1 Qu'est ce que la quantification des incertitudes ?

Dans de nombreux domaines d'applications, on est confronté à des incertitudes sur les données, par exemple certaines propriétés physiques d'un matériau, d'un milieu. Ces incertitudes peuvent provenir d'imprécisions sur les mesures ou d'un nombre de mesures insuffisant, typiquement dans le cas d'un milieu très hétérogène pour lequel de très nombreuses mesures seraient nécessaires. On doit alors inclure ces incertitudes dans les modèles. On peut pour cela modéliser ces données par des champs aléatoires. On est alors amené à résoudre des équations aux dérivées partielles à coefficients aléatoires. L'inconnue est alors un champ aléatoire, dont on souhaite connaître la loi. En pratique, on se contente souvent de l'espérance et de l'écart-type. On souhaite connaître comment les incertitudes sur les données se propagent sur la solution. Ce problème de quantification des incertitudes apparaît dans de nombreux domaines tels que l'étude de la déformation de matériaux inhomogènes (bois, mousses, biomatériaux), l'étude de l'action du vent sur des structures, la sismologie, l'étude des matériaux composites, les vibrations aléatoires, la gestion des ressources en pétrole ; voir par exemple [3, 19, 20, 31, 45, 61] et les références qui s'y trouvent. Les méthodes et résultats que l'on présente ici ont une portée générale, néanmoins ce travail est guidé par des applications en hydrogélogie que l'on va maintenant préciser.

1.1.2 Un problème de quantification des incertitudes en hydrogéologie

L' application qui sous-tend ce travail est la modélisation numérique des écoulements dans les nappes phréatiques et de la manière dont les polluants s'y répandent. Nous présentons donc ici cette application.

Ecoulement

On considère donc un milieu poreux saturé, dans lequel s'écoule un fluide monophasique incompressible, en régime stationnaire. De manière classique, on utilise pour modéliser cet écoulement la loi de Darcy

$$v(x) = -a(x)\nabla u(x),$$

où u est la pression hydraulique, a le tenseur de perméabilité et v la vitesse de Darcy. Pour simplifier, on supposera souvent le tenseur de perméabilité scalaire, ce qui correspond à considérer un milieu poreux isotrope. La perméabilité a traduit la facilité avec laquelle l'eau peut s'écouler pour un gradient de pression donné. On complète cette équation avec l'équation d'incompressibilité $\mathrm{div}(v(x)) = 0$, pour obtenir l'équation d'écoulement usuelle

$$-\mathrm{div}(a(x)\nabla u(x)) = 0, \tag{1.1}$$

qu'on complète avec des conditions aux bords. Souvent, les formations géologiques naturelles présentent de fortes hétérogénéités, et il n'est en pratique possible de mesurer la perméabilité qu'en un nombre très limité

d'endroits. Ceci est la principale source d'incertitudes dans le calcul des écoulements souterrains. Pour modéliser ces incertitudes, des modèles stochastiques ont été proposés dès les années 80 par des hydrogéologues, voir [16, 17, 14, 28] par exemple. Le champ de perméabilité a devient alors un champ aléatoire, c'est-à-dire une fonction de la position x et d'un paramètre aléatoire ω, et on remplace alors l'équation déterministe (1.1) par l'équation aux dérivées partielles à coefficients aléatoires : pour presque tout ω

$$-\mathrm{div}(a(\omega,x)\nabla u(\omega,x)) = 0, \tag{1.2}$$

que l'on complète à nouveau par des conditions aux bords. Un modèle très fréquemment utilisé par les hydrogéologues est de prendre pour a un champ lognormal homogène, c'est-à-dire $a(\omega,x) = e^{g(\omega,x)}$, où g est un champ gaussien, dont la loi est caractérisée par son espérance et sa fonction de covariance $\mathrm{cov}[g](x,y) = \mathbb{E}[(g(\omega,x)-\mathbb{E}[g(\omega,x)])(g(\omega,y)-\mathbb{E}[g(\omega,y)])]$. On suppose que $\mathrm{cov}[g](x,y)$ ne dépend que de la distance de x à y, c'est-à-dire que le champ est homogène. Ce modèle simple garantit que l'équation (1.2) admet une unique solution presque partout, et il donne des réalisations de a qui varient sur plusieurs ordres de grandeur, ce qui est typique dans les écoulements souterrains. Dans le cadre des applications, un choix fréquent de fonction de covariance est la covariance exponentielle [25, 42]

$$\mathrm{cov}[g](x,y) = \sigma^2 e^{-\frac{\|x-y\|}{\ell}}, \tag{1.3}$$

où σ est l'écart-type et l la longueur de corrélation. Comme nous le verrons en détail par la suite, cette fonction de covariance étant peu régulière, les réalisations du champ de perméabilité associé sont également peu régulières, ce qui en fait un bon modèle car dans les formations géologiques naturelles, le champ de perméablité a tendance à être peu régulier à cause des processus de sédimentation. Des fonctions de covariance plus générales de la forme

$$\mathrm{cov}[g](x,y) = \sigma^2 e^{-\left(\frac{\|x-y\|}{\ell}\right)^{\delta}} \tag{1.4}$$

sont utilisées. Dans le cadre des applications considérées ici, on s'intéresse esentiellement au cas où $\sigma^2 \geq 1$, qui correspond à des incertitudes importantes et où $\ell \leq 1$, qui correspond à une longueur de corrélation inférieure à la taille de la zone à laquelle on s'intéresse, voire même $\ell \ll 1$.

Pour conclure ce paragraphe, il est important d'évoquer dès maintenant deux particularités d'un tel choix de champ de perméabilité, qu'on reverra plus en détail ultérieurement et qui excluent l'application de nombreux résultats qu'on peut trouver dans la littérature. Tout d'abord un champ lognormal n'est clairement ni uniformément borné par rapport à ω, ni uniformément coercif par rapport à ω. De plus, avec le choix de la covariance exponentielle (1.3), les trajectoires de a sont peu régulières, elles ne sont en particulier pas $\mathcal{C}^1$ ou $W^{1,\infty}$. Pour plus de détails sur les propriétés du champ décrit ci-dessus, on se reportera au paragraphe 1.2.1, et pour plus de détails sur les résultats qu'on peut trouver dans la littérature, on se reportera au paragraphe 1.1.3 et aux références qui s'y trouvent.

Advection-diffusion de polluants

On s'intéresse maintenant à l'advection-diffusion d'un soluté (typiquement un polluant) dans le milieu décrit ci-dessus. Plus précisément, on considère un soluté inerte injecté à l'instant initial dans le milieu poreux décrit ci-dessus et on suppose qu'il est transporté par l'écoulement de l'eau et diffusé. On considère uniquement la diffusion moléculaire, négligeant la dispersion cinématique. On suppose que la diffusion moléculaire est homogène et isotrope et que la porosité est constante, égale à 1. Ce type de migration d'un soluté est décrit par l'équation d'advection-diffusion suivante :

$$\frac{\partial c(\omega,x,t)}{\partial t} + v(\omega,x).\nabla c(\omega,x,t) - D\Delta c(\omega,x,t) = 0, \tag{1.5}$$

où D est le coefficient de diffusion moléculaire, v la vitesse de Darcy définie dans le paragraphe précédent, et c la concentration de soluté. On s'intéresse au cas où l'advection est dominante, c'est-à-dire au cas où le nombre de Péclet $\frac{\ell\|v\|_{mean}}{D}$ est important (typiquement ≥ 100). La condition initiale à $t = 0$ est l'injection du soluté, c'est-à-dire $c(t=0) = \mathbb{1}_R$, où R est un rectangle inclus dans le domaine O. L'équation (1.5) doit

être complétée avec des conditions aux bords. D'un point de vue applicatif, on souhaite savoir comment un polluant injecté à un instant donné de manière localisée dans une nappe phréatique va se propager dans celle-ci, on peut en particulier se demander quelle taille va atteindre la zone polluée, à quelle vitesse elle va se déplacer, combien de temps la pollution va mettre pour atteindre une zone de pompage.

On définit maintenant les deux quantités que l'on souhaite ici calculer : l'extension moyenne et la (macro)-dispersion moyenne. Pour cela on commence par définir le centre de masse du soluté :

$$G(\omega,t)=\int_O c(\omega,x,t)xdx.$$

L'extension du soluté $S(\omega,t)$ est alors définie par

$$S(\omega,t)=\int_O c(\omega,x,t)(x-G(\omega,t))(x-G(\omega,t))^t dx,$$

et la dispersion $\mathcal{D}(\omega,t)$ comme la dérivée temporelle de l'extension :

$$\mathcal{D}(\omega,t)=\frac{dS(\omega,t)}{dt}.$$

On définit enfin l'extension moyenne et la dispersion moyenne.

$$S(t)=\mathbb{E}[S(\omega,t)] \quad \text{et} \quad \mathcal{D}(t)=\mathbb{E}[\mathcal{D}(\omega,t)].$$

La macro-dispersion traduit la vitesse à laquelle le panache de soluté s'étend. La détermination du coefficient de macro-dispersion, et en particulier de l'influence de l'hétérogénéité du milieu sur ce coefficient, a été le sujet de nombreux travaux ces 25 dernières années, voir par exemple [5, 14, 18, 28, 65, 67, 62]. Néanmoins, dans la plupart de ces travaux, la diffusion a été négligée, considérant uniquement le terme d'advection. Par conséquent, on est entre autres intéressés par l'influence qualitative et quantitative de la diffusion moléculaire sur la macro-dispersion (voir [15] pour des résultats numériques).

1.1.3 Méthodes et estimations d'erreurs connues pour la quantification des incertitudes

Dans cette partie, on présente plusieurs des principales méthodes numériques utilisées pour la quantification des incertitudes, ainsi qu'un outil important d'approximation dans un espace stochastique de dimension finie : le développement de Karhunen-Loève, cette approximation jouant entre autres un rôle fondamental dans les méthodes spectrales stochastiques. Pour simplifier cet exposé, on présente ces méthodes dans le cas de l'équation modèle suivante : pour presque tout $\omega\in\Omega$

$$\left\{\begin{array}{rcll} -\mathrm{div}(a(\omega,x)\nabla u(\omega,x)) & = & f(x) & x\in D,\\ u(\omega,x) & = & 0 & x\in\partial D,\end{array}\right. \tag{1.6}$$

néanmoins les méthodes décrites ci-dessous ont une portée plus générale, certaines plus que d'autres. Afin que le problème soit bien posé, on suppose dans ce paragraphe que $f\in L^2(D)$, et que pour presque tout ω, $x\mapsto a(\omega,x)$ est mesurable, et qu'il existe $a^{min}(\omega)>0$ et $a^{max}(\omega)<+\infty$ tels que pour presque tout $x\in D$ on a $a^{min}(\omega)\leq a(\omega,x)\leq a^{max}(\omega)$. On suppose que D est un ouvert borné convexe inclus dans $\mathbb{R}^d$. On considère V_h un espace standard d'éléments finis sur D de fonctions continues, linéaires par morceaux, associé à une famille régulière de triangulations de D dont on note h le diamètre maximum du maillage. Pour alléger l'exposé, on ne donnera que les estimations d'erreur H^1, mais bien sûr dans chaque cas on a également des estimations d'erreur L^2.

Une hypothèse faite par de nombreux auteurs pour obtenir des estimations d'erreurs est la suivante :

Hypothèse 1.1.1. *On suppose que $a\in L^\infty(\Omega,\mathcal{C}^1(\bar{D}))$ et qu'il existe $a^{min}>0$ tel que pour presque tout ω on ait $a^{min}\leq a(\omega,x)$ pour tout $x\in\bar{D}$.*

Cette hypothèse contient à la fois une hypothèse de régularité des réalisations de a (pour appliquer la théorie standard d'estimations d'erreur éléments finis) et une hypothèse d'existence d'une borne et d'une constante de coercivité uniforme par rapport à ω (pour pouvoir obtenir des estimations d'erreurs uniformes par rapport à ω). On notera dès à présent que les champs lognormaux définis au paragraphe "Ecoulement" de la section 1.1.2 ne vérifient pas le deuxième point, et que dans le cas d'une covariance exponentielle (1.3), le premier point n'est pas vérifié non plus.

Méthodes de type Monte-Carlo

Le principe de la méthode de Monte-Carlo est simple, on considère M réalisations indépendantes du champ de perméabilité $a^1, ...a^M$, pour chacune de ces réalisations on calcule une approximation u_h^i de la solution u^i de l'EDP déterministe correspondante

$$-\mathrm{div}(a^i(x)\nabla u^i(x)) = f(x),$$

dans l'espace d'éléments finis V_h (on peut utiliser n'importe quelle autre méthode numérique pour calculer la solution de l'EDP déterministe, on ne considèrera ici que des méthode d'éléments finis mais le principe est tout à fait général). Pour une fonction $\varphi : \mathbb{R} \to \mathbb{R}$ ou $\varphi : H^1(D) \to \mathbb{R}$, on approche alors $\mathbb{E}[\varphi(u)]$ par $\frac{1}{M}\sum_{i=1}^{M} \varphi(u_h^i)$. On peut alors décomposer l'erreur commise en un terme correspondant à la discrétisation spatiale et un terme correspondant au Monte-Carlo, qui est en $1/\sqrt{M}$. Plus précisément, on a par exemple le résultat suivant [2] pour le cas du calcul de l'espérance.

Proposition 1.1.2. *On suppose l'hypothèse 1.1.1 vérifiée, on a alors l'estimation d'erreur suivante : il existe une constante C telle que pour tout $h > 0$ et tout M on ait*

$$\left\| \mathbb{E}[u] - \frac{1}{M}\sum_{i=1}^{M} u_h^i \right\|_{L^2(\Omega, H_0^1(D))} \leq C\left(h + \frac{1}{\sqrt{M}}\right) \|f\|_{L^2(D)}. \tag{1.7}$$

Les méthodes de type Monte-Carlo ont l'avantage d'être faciles à mettre en oeuvre et parallélisables, ainsi que d'avoir un cadre d'application très général : elles peuvent s'appliquer avec n'importe quelle EDP pour laquelle on dispose d'une méthode numérique (déterministe), que l'on peut alors utiliser en boîte noire. La vitesse de convergence est de plus indépendante de la "dimension stochastique de a" (voir paragraphe "Méthodes spectrales stochastiques" ci-dessous). Néanmoins elles ont pour inconvénient d'avoir une vitesse de convergence assez lente ($1/\sqrt{M}$), le côut de calcul étant M fois celui du problème déterministe (à savoir celui de la résolution d'un système linéaire de taille $\dim(V_h)$).

Des améliorations de cette méthode de base sont possible, en particulier grâce à des techniques de réduction de variance, ou en utilisant des méthodes de type quasi Monte-Carlo [38].

Un outil d'approximation dans un espace stochastique de dimension finie : le développement de Karhunen-Loève

Le point de départ des méthodes spectrales stochastiques qu'on va présenter dans le paragraphe "Méthodes spectrales stochastiques" ci-dessous, qui est un point crucial, est l'approximation du coefficient a dans un espace aléatoire de dimension finie, c'est-à-dire par une fonction d'un nombre fini de variables aléatoires. Les deux méthodes fréquemment utilisées pour obtenir une telle approximation sont les développements en polynômes de Chaos et les développements de Karhunen-Loève. On expose ici cette dernière méthode [2, 31] qui est celle sur laquelle on se concentrera par la suite, car elle est bien adaptée au cas d'un champ logonormal homogène exposé dans le paragraphe 1.1.2, pour des raisons que l'on expliquera. Le développement de Karhunen-Loève est également utilisé pour certaines méthodes de type petites perturbations (voir paragraphe "Méthodes de type petites perturbations"). C'est enfin une des méthodes utilisées pour simuler le champ a, dans le cadre des méthodes de type Monte-Carlo par exemple.

Soit $b : \Omega \times D \to \mathbb{R}$ un processus stochastique dont la fonction de covariance $\mathrm{cov}[b]$ est continue sur $\bar{D} \times \bar{D}$. On considère l'opérateur de Hilbert-Schmidt auto-adjoint positif associé défini par

$$h \in L^2(D) \mapsto \int_D \mathrm{cov}[b](x, \cdot)h(x)dx.$$

On note $(\lambda_n, b_n)_{n\in\mathbb{N}^*}$ la suite de couples propres de cet opérateur, où les valeurs propres λ_n sont rangées dans l'ordre décroissant. On rappelle qu'elles vérifient $\sum_{n\geq 1} \lambda_n = \int_D \mathrm{cov}[b](x, x)dx$ et que les vecteurs propres b_n forment une famille orthonormale. On a alors

$$b(\omega, x) \overset{L^2(\Omega\times D)}{=} \mathbb{E}[b](x) + \sum_{n\geq 1} \sqrt{\lambda_n} b_n(x) Y_n(\omega),$$

où les variables aléatoires Y_n sont centrées réduites et décorrélées (et donc forment une famille orthonormale), définies pour $\lambda_n > 0$ de manière unique par

$$Y_n(\omega) = \frac{1}{\sqrt{\lambda_n}} \int_D (b(\omega, x) - \mathbb{E}[b](x))b_n(x)dx.$$

D'après le théorème de Mercer [60], si on définit la somme partielle $b_N(\omega, x) = \mathbb{E}[b](x) + \sum_{n\geq 1}^{N} \sqrt{\lambda_n} b_n(x) Y_n(\omega)$, on a

$$\sup_{x\in D} \mathbb{E}[(b - b_N)^2](x) \underset{N\to+\infty}{\to} 0. \tag{1.8}$$

Il est important de noter que dans le cas où b est un champ gaussien, on a des propriétés supplémentaires très intéressantes. En effet si b est un champ gaussien, on connaît la loi des Y_n : ce sont des gaussiennes centrées réduites, ce qui est très utile en pratique, de plus dans ce cas les variables aléatoires Y_n sont indépendantes, et pas seulement décorrélées. De plus, dans le cas d'un champ gaussien avec corrélation exponentielle (définie par (1.3)), on a des expressions quasi-explicites pour les valeurs propres dans le cas où on prend la norme 1 et se place sur un domaine rectangulaire (pour un domaine quelconque, on peut utiliser le développement de Karhunen-Loève sur un domaine rectangulaire contenant ce domaine). En effet, en dimension 1, considérant l'équation caractéristique

$$(\ell^2 w^2 - 1)\sin(w) = 2\ell w \cos(w)$$

et notant $(w_n)_{n\geq 1}$ la suite de ses racines positives ordonnées dans l'ordre croissant, alors les valeurs propres du développement de Karhunen-Loève de b sont définies par

$$\lambda_n = \frac{2\ell\sigma^2}{\ell^2 w_n^2 + 1} \tag{1.9}$$

et les vecteurs propres par

$$b_n(x) = \alpha_n(\sin(w_n x) + \ell w_n \cos(w_n x)), \tag{1.10}$$

où $\alpha_n = \frac{1}{\sqrt{(\ell^2 w_n^2+1)/2+\ell}}$; pour la preuve, voir [31, 76] par exemple. En dimension supérieure, les couples propres sont obtenus en tensorisant les couples propres mono-dimensionnels. Des méthodes numériques efficaces ont été développées pour calculer les couples propres du développement de Karhuen-Loève dans le cas d'une covariance quelconque, voir par exemple [64, 22].

On utilise finalement le développement de Karhunen-Loève pour approcher le processus stochastique b dans un espace stochastique de dimension finie, c'est-à-dire par une fonction d'un nombre fini de variables aléatoires, en l'occurence par $b_N(\omega, x) = \mathbb{E}[b](x) + \sum_{n=1}^{N} \sqrt{\lambda_n} b_n(x) Y_n(\omega)$. On peut utiliser le développement de Karhunen-Loève pour le coefficient a de l'EDP (1.6), ou dans le cas d'un champ lognormal vu au paragraphe "Ecoulement" de la section 1.1.2 pour le champ gaussien $g = \log(a)$. Dans ce cas, le développement de

Karhunen-Loève est un outil bien adapté, car g étant un champ gaussien, son développement de Karhunen-Loève fait apparaître des variables aléatoires gaussiennes centrées réduites indépendantes, faciles à simuler, et de plus, dans le cas de la covariance exponentielle (1.3), les fonctions propres sont faciles à calculer. On approche donc le coefficient lognormal $a = e^g$ par

$$a_N(\omega, x) = e^{g_N(\omega,x)}, \quad \text{où} \quad g_N(\omega, x) = \mathbb{E}[g](x) + \sum_{n=1}^{N} \sqrt{\lambda_n} b_n(x) Y_n(\omega), \tag{1.11}$$

où $(\lambda_n, b_n)_{n\geq 1}$ est la suite des couples propres de l'opérateur de Hilbert-Schimdt de noyau la fonction de corrélation de $g = \log(a)$.

Méthodes spectrales stochastiques

Les méthodes spectrales stochastiques, qui comprennent les méthodes de Galerkin stochastiques [2, 46, 31, 54, 24, 72] et les méthodes de collocation stochastiques [1, 75, 73] ont suscité beaucoup d'intérêt ces dernières années. Ces méthodes s'appliquent pour des coefficients a de dimension stochastique finie, on commence donc par approcher le champ a par une approximation a_N, fonction de N variables aléatoires, c'est-à-dire dans un espace stochastique de dimension N. Comme on vient de le voir, le développement de Karhunen-Loève permet d'obtenir une telle approximation (les polynômes de Chaos sont également fréquemment utilisés, voir par exemple [31, 72, 30, 29, 72, 46, 47]). Les méthodes spectrales stochastiques permettent alors de calculer la solution u_N de l'équation "approchée" suivante :

$$\left\{ \begin{array}{rcll} -\mathrm{div}(a_N(\omega, x)\nabla u_N(\omega, x)) & = & f(x) & x \in D, \\ u_N(\omega, x) & = & 0 & x \in \partial D. \end{array} \right. \tag{1.12}$$

Une question importante est alors celle de la quantification de l'erreur commise en approchant u par u_N. Cette question est d'autant plus importante, que comme nous allons le voir, le coût des méthodes spectrales stochastiques croît de manière importante avec la dimension stochastique du coefficient N (cette croissance est typiquement exponentielle pour les versions basiques des méthodes spectrales stochastiques, voir ci-dessous). Une première réponse à cette question est donnée dans [2] sous forme d'estimation d'erreur forte, sous des hypothèses très fortes (de décroissance des valeurs propres et d'existence d'une borne uniforme pour les Y_n) et via une estimation $L^\infty(\Omega \times D)$ de l'erreur commise sur a, en particulier ce résultat ne s'applique pas pour un champ lognormal.

Nous présentons maintenant les méthodes spectrales stochastiques à proprement parler. a_N étant une fonction de N variables aléatoires $(Y_1, ...Y_N)$, on utilisera la notation $a_N(\omega, x) = \tilde{a}_N(Y_1(\omega), ..., Y_N(\omega), x)$. On suppose que le vecteur aléatoire $Y = (Y_1, ..., Y_N)$ admet une densité (jointe) ρ par rapport à la mesure de Lebesgue sur $\mathbb{R}^N$ (c'est par exemple le cas pour un champ lognormal), on va pouvoir ainsi ramener la famille d'EDP (1.12) paramétrée par ω à une famille d'EDP paramétrée dans $\mathbb{R}^N$. En effet, si pour y appartenant au support $\Gamma \subset \mathbb{R}^N$ de ρ on note $\tilde{u}_N(y, \cdot)$ la solution de l'EDP

$$\left\{ \begin{array}{rcll} -\mathrm{div}(\tilde{a}_N(y, x)\nabla \tilde{u}_N(y, x)) & = & f(x) & x \in D, \\ \tilde{u}_N(y, x) & = & 0 & x \in \partial D, \end{array} \right. \tag{1.13}$$

on a alors $u_N(\omega, x) = \tilde{u}_N(Y(\omega), x)$.

Les méthodes de Garlerkin stochastiques se basent alors sur le fait que le problème de départ (1.12) est équivalent à la formulation variationnelle globale suivante : $\tilde{u}_N$ est la fonction appartenant à $L^2_\rho(\Gamma, H^1_0(D))$ vérifiant pour tout $v \in L^2_\rho(\Gamma, H^1_0(D))$

$$\int_\Gamma \int_D \tilde{a}_N(y, x)\nabla \tilde{u}_N(y, x)\nabla v(y, x)\rho(y)dxdy = \int_\Gamma \int_D f(x)v(y, x)\rho(y)dxdy. \tag{1.14}$$

En se basant sur cette reformulation, on approche u_N dans l'espace de dimension finie $P \otimes V_h$ par une méthode de Galerkin, où P est un espace de polynômes ou de polynômes par morceaux en N variables et V_h l'espace d'éléments finis défini précédemment. Dans le cas où P est l'espace des polynômes de degré inférieur ou égal à p (p étant un multi-indice de longueur N), on note $\tilde{u}_N^{h,p}$ la solution approchée par projection de

Galerkin, et on définit $u_N^{h,p}(\omega,x) = \tilde{u}_N^{h,p}(Y(\omega),x)$. Notant P_p l'espace des polynômes en N variables de degré p, $\tilde{u}_N^{h,p}$ est défini comme l'unique fonction de $P_p \otimes V_h$ vérifiant pour tout $v \in P_p \otimes V_h$

$$\int_\Gamma \int_D \tilde{a}_N(y,x)\nabla \tilde{u}_N^{h,p}(y,x)\nabla v(y,x)\rho(y)dxdy = \int_\Gamma \int_D f(x)v(y,x)\rho(y)dxdy. \tag{1.15}$$

Sous des hypothèses légèrement plus fortes que l'Hypothèse 1.1.1, que l'on ne détaillera pas ici, l'estimation d'erreur suivante est prouvée dans [2] :

Théorème 1.1.3. *Il existe une constante $C > 0$ telle que pour tout $\tau \in]0,1[$, il existe $(r_n)_n \in]0,1[^N$ tels que pour tout $h > 0$ et $p \in \mathbb{N}^N$ on ait*

$$\|\mathbb{E}[u_N - u_N^{h,p}]\|_{H^1(D)} \leq C\left(h + \frac{1}{\tau}\sum_{n=1}^N (r_n)^{p_n+1}\right).$$

Les méthodes de collocation stochastiques se basent quant à elles sur la reformulation de (1.12) sous forme d'une famille d'EDP paramétrée par $y \in \mathbb{R}^N$, et sur le fait que pour presque tout $x \in D$, la loi de $\omega \in \Omega \mapsto u_N(\omega,x)$ est celle de $y \in \Gamma \mapsto \tilde{u}_N(y,x)$, où on a muni Γ de la mesure ayant pour densité ρ par rapport à la mesure de Lebesgue. On est donc ramené à un problème d'intégration numérique sur $\mathbb{R}^N$ avec poids ρ. Le principe général des méthodes de collocation stochastiques est de calculer une solution approchée $\tilde{u}_N^h(y,\cdot)$ dans l'espace d'éléments finis V_h de la solution $\tilde{u}_N(y,\cdot)$ de (1.12) pour certaines valeurs de y : $y_1, ...y_{N_p}$, appelés points de collocation. On en déduit alors une solution approchée $\tilde{u}_N^{h,p}$, à partir des valeurs obtenues en les points de collocation en utilisant une interpolation de Lagrange ou des polynômes de Chaos généralisés.

On détaille ici la première possibilité, suivant la méthode proposée dans [1]. On suppose ici les variables aléatoires $(Y_n)_{n=1...N}$ indépendantes, on peut donc écrire ρ sous la forme $\rho(y) = \rho_1(y_1)...\rho_N(y_N)$. Soit p un multi-indice de longueur N, pour $1 \leq n \leq N$, on note $(y_{n,k_n})_{1\leq k_n \leq p_n+1}$ l'ensemble des p_n+1 racines du pôlynome orthogonal de degré p_n+1 associé au poids ρ_n. On tensorise ensuite les points de quadrature choisis dans chaque direction. En associant à tout vecteur d'indices $[k_1,...,k_N]$ un indice global k, on obtient un ensemble de points d'interpolation $y_k = [y_{1,k_1},...,y_{N,k_N}]$ pour $1 \leq k \leq N_p$. Ces points sont les noeuds de la méthode de quadrature associée au poids ρ. On approche finalement $\tilde{u}_N(y,\cdot)$ par l'interpolée de Lagrange $\tilde{u}_N^{h,p}$ de $y \mapsto \tilde{u}_N^h(y,\cdot)$ aux points $(y_k)_{k=1...N_p}$, et on définit naturellement $u_N^{h,p}(\omega,x) = \tilde{u}_N^{h,p}(Y(\omega),x)$.

Sous l'hypothèse 1.1.1, l'estimation d'erreur suivante est prouvée dans [1] :

Théorème 1.1.4. *Il existe des constantes positives $r_1,...,r_N$ et une constante C telles que pour tout $h > 0$ et pour tout $p \in \mathbb{N}^N$*

$$\|\mathbb{E}[u_N - u_N^{h,p}]\|_{H^1(D)} \leq C\left(h + \sum_{n=1}^N \sqrt{p_n}e^{-r_n\sqrt{p_n}}\right). \tag{1.16}$$

En fait, on trouve dans [1] une estimation d'erreur en norme $L^2(\Omega, H^1(D))$ dont le résultat ci-dessus est une conséquence, ainsi qu'une meilleure estimation d'erreur dans le cas où Γ est borné. De nombreuses améliorations de cette méthode de collocation ont été proposées permettant de réduire très nettement le coût de calcul, telles que l'utilisation de grilles de points de collocation creuses ; voir par exemple [59, 58].

Dans le cas des méthodes de Galerkin stochastiques, comme dans le cas des méthodes de collocation stochastiques, on obtient une convergence exponentielle par rapport au degré du polynôme dans chaque direction, et donc très rapide. Dans les deux cas (Théorème 1.1.3 et Théorème 1.1.4), la preuve est basée sur un résutat d'analycité de $y \in \mathbb{R}^N \mapsto u(y,\cdot) \in H_0^1(D)$. Le coût de calcul de la méthode de Galerkin stochastique est celui de la résolution d'un système linéaire de taille $\dim(P_p)\dim(V_h)$, soit $(1+q)^N\dim(V_h)$ dans le cas où on prend des pôlynomes de même degré q par rapport à chaque variable, le coût de calcul de la méthode de collocation stochastique est quant à lui celui de la résolution de N_p systèmes linéaires de taille $\dim(V_h)$, avec $N_p = q^N$ dans le cas où on prend le même nombre de points q dans chaque direction. Dans les deux cas, ce coût augmente de manière importante avec N. Ce coût est moindre dans les améliorations de ces méthodes, mais reste important pour N grand. Par conséquent, ces méthodes sont très efficaces pour des valeurs de N relativement faibles, mais inapplicables quand N devient trop important. La question

de l'approximation de a dans un espace stochastique de dimension finie est donc une question cruciale. On notera d'une part que les méthodes de collocation stochastiques s'étendent naturellement à une grande classe d'EDP, y compris non linéaires, alors que c'est moins évident pour les méthodes de Galerkin stochastiques. D'autre part les méthodes de collocation stochastiques permettent d'utiliser des codes préexistants pour l'EDP déterministe, contrairement aux méthodes de Galerkin stochastiques.

Méthodes de type petites perturbations

Les méthodes de type petites perturbations consistent à développer le champ aléatoire en série (de Taylor ou de Karhunen-Loève) autour de sa moyenne, et de tronquer cette série à un certain ordre ; voir par exemple [45, 48, 51]. Ces méthodes sont donc réservées aux cas où on a de petites incertitudes.

On va maintenant présenter les principaux résultats de cette thèse, dans une première partie on s'intéresse à des méthodes numériques pour l'équation d'écoulement présentée dans le paragraphe "Ecoulement" de la section 1.1.2, dans une seconde partie on s'intéresse à des méthodes numériques pour le couplage de l'équation d'écoulement et de l'équation d'advection-diffusion présenté dans le paragraphe Advection-diffusion de polluants de la section 1.1.2.

1.2 Analyse numérique de l'équation d'écoulement

Dans cette section on s'intéresse à l'analyse numérique de différentes méthodes pour l'équation d'écoulement (1.2) avec coefficient lognormal, présentée dans le paragraphe "Ecoulement" de la section 1.1.2. Les résultats de cette section sont détaillés dans les Chapitres 2, 3 et 4. Avant de s'intéresser à l'analyse numérique, on commence par définir un cadre précis, et par donner deux résultats préliminaires qui seront très utiles par la suite, et qui permettent en particulier d'obtenir un résultat d'existence et d'unicité de solution dans $L^p(\Omega, H_0^1(D))$. Les résultats de cette partie sont détaillés dans les Chapitres 2, 3 et 4.

1.2.1 Présentation du problème et résultats préliminaires

Soit D un domaine borné de $\mathbb{R}^d$ de classe $\mathcal{C}^2$ et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Soit $f \in L^2(D)$ et a un champ lognormal homogène sur $\bar{D}$, c'est-à-dire $a(\omega, x) = e^{g(\omega,x)}$, où g est un champ gaussien homogène, de moyenne supposée nulle pour simplifier l'exposé, et de covariance

$$\mathrm{cov}[g](x, y) = k(\|x - y\|), \tag{1.17}$$

où $k \in \mathcal{C}^{0,1}(\mathbb{R}, \mathbb{R})$. On s'intéresse à l'équation aux dérivées partielles elliptique à coefficients aléatoires suivante : pour presque tout ω dans Ω

$$\left\{ \begin{array}{rcll} -\mathrm{div}(a(\omega, x)\nabla u(\omega, x)) & = & f(x) & x \in D, \\ u(\omega, x) & = & 0 & x \in \partial D. \end{array} \right. \tag{1.18}$$

On obtient donc une famille d'équations aux dérivées partielles paramétrée par un paramètre aléatoire ω. C'est pourquoi, dans ce qui suit, pour montrer des propriéts d'existence et d'unicité et plus tard d'approximation spatiale, on raisonnera sur le problème déterministe à ω fixé, puis on s'intéressera à expliciter la dépendance en ω. On considère un second membre déterministe pour simplifier l'exposé, mais les résultats qui suivent peuvent être facilement étendus au cas où f est également aléatoire, sous les hypothèses adéquates. Pour commencer, on donne deux résultats préliminaires fondamentaux pour la suite. On a tout d'abord un résultat de régularité des réalisations de g et donc de a, conséquence du théorème de Kolmogorov.

Proposition 1.2.1. *g admet une version dont les trajectoires sont α-hölderiennes, pour tout $\alpha < 1/2$.*

Dans ce qui suit, on identifie g à cette version. On en déduit en particulier qu'il existe pour presque tout ω, par continuité, des constantes $a^{min}(\omega) > 0$ et $a^{max}(\omega) < +\infty$ telles que $a^{min}(\omega) \leq a(\omega, x) \leq a^{max}(\omega)$ pour tout $x \in D$. L'équation aux dérivées partielles (1.18) admet donc une unique solution pour presque

tout ω d'après le théorème de Lax-Milgram. Le champ a étant lognormal, il n'est ni uniformément borné ni coercif par rapport à ω, néanmoins, on déduit du théorème de Fernique la propriété un peu plus faible suivante :

Proposition 1.2.2. *$\frac{1}{a^{min}(\omega)} \in L^p(\Omega)$ et $a^{max}(\omega) \in L^p(\Omega)$ pour tout $p > 0$.*

Ce résultat, ainsi que des résultats similaires pour des approximations de a que l'on verra par la suite, qui proviennent d'une application du théorème de Fernique, nous seront très utiles pour obtenir des bornes L^p pour la solution et par la suite pour les estimations d'erreur d'éléments finis, ayant au préalable obtenu des majorations à ω fixé, exprimées en fonctions de $a^{max}(\omega)$ et $1/a^{min}(\omega)$. Suivant cette démarche, on déduit pour commencer le résultat d'existence et d'unicité suivant.

Proposition 1.2.3. *L'équation* (1.18) *admet une unique solution u, qui appartient à $L^p(\Omega, H_0^1(D))$, pour tout $p > 0$.*

On notera que le caractère bien posé et la question de l'approximation numérique pour une EDP elliptique à coefficient lognormal ont été étudiés indépendamment dans [36], dans [26] avec une approche par bruit blanc et récemment dans [66], qui traite du problème inverse.

1.2.2 Erreur de troncature du développement de Karhunen-Loève

Comme on l'a vu dans le paragraphe "Méthodes spectrales stochastiques" de la section 1.1.3, la question de l'approximation du coefficient a dans un espace de faible dimension stochastique est cruciale pour les méthodes spectrales stochastiques. On va donc chercher à quantifier l'erreur commise sur la solution de l'EDP en approchant a par a_N, où a_N est obtenu en tronquant le développement de Karhunen-Loève de $\log(a)$ à l'ordre N, défini en (1.11). Rappelant que la solution de l'EDP (1.18) avec coefficient a est notée u, et celle de l'EDP avec le coefficient approché a_N (1.12) est noté u_N, on commence par donner une estimation d'erreur forte (c'est-à-dire qu'on compare $u_N(\omega, x)$ et $u(\omega, x)$ pour "le même ω") puis on donne une estimation d'erreur faible (c'est-à-dire qu'on compare la loi de u_N et la loi de u, en estimant des quantités du type $\mathbb{E}[\varphi(u_N)] - \mathbb{E}[\varphi(u)]$). Pour obtenir ces estimations on va faire l'hypothèse suivante sur les couples propres associés au développement de Karhunen-Loève de $g = \log(a)$.

Hypothèse 1.2.4. *On suppose que*

i) les fonctions propres b_n appartiennent à $\mathcal{C}^1(\bar{D})$,

ii) la série

$$\sum_{n\geq 0} \lambda_n \|b_n\|_\infty^2$$

est convergente,

iii) qu'il existe $0 < \alpha < 1$ tel que la série

$$\sum_{n\geq 0} \lambda_n \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha}$$

soit convergente.

On introduit alors la notation suivante :

Définition 1.2.5. *Pour $0 \leq \alpha < 1$ vérifiant la condition de l'Hypothèse 1.2.4 et $N \in \mathbb{N}$, on note*

$$R_N^\alpha = \max\left(\sum_{n>N} \lambda_n \|b_n\|_\infty^2, \sum_{n>N} \lambda_n \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha}\right).$$

Pour obtenir de la convergence presque sûre, on fait l'hypothèse suivante.

Hypothèse 1.2.6. *Soit $\alpha > 0$ comme dans l'Hypothèse 1.2.4, on suppose qu'il existe $p_0 > 0$ tel que la série*

$$\sum_{N \geq 0} (R_N^\alpha)^{p_0}$$

soit convergente.

Sous les hypothèses 1.2.4 et 1.2.6, on peut montrer le résultat d'erreur forte suivant.

Théorème 1.2.7. *$(u_N)_N$ converge presque sûrement vers u dans $H_0^1(D)$ et on a l'estimation d'erreur suivante : pour tout $p > 0$ et α comme dans l'Hypothèse 1.2.4, il existe une constante $C_{1.2.7}(\alpha, p)$ telle que*

$$\|u - u_N\|_{L^p(\Omega, H_0^1(D))} \leq C_{1.2.7}(\alpha, p)\sqrt{R_N^\alpha}. \tag{1.19}$$

On remarquera que cette estimation nous intéressera souvent pour α proche de 0. En effet, dans le cas où les b_n sont uniformément bornés l'erreur forte est presque de l'ordre de la racine du reste de la série des valeurs propres.

Pour montrer ce résultat, on commence par obtenir une estimation d'erreur forte sur les coefficients, c'est-à-dire par majorer $\|a - a_N\|_{L^p(\Omega, \mathcal{C}^0(\bar{D}))}$. Pour ce faire on commence par estimer $g_N - g$ en utilisant le caractère gaussien et le thérorème de Kolmogorov comme pour la preuve de la Proposition 1.2.1, mais cette fois-ci on estime la norme dans l'espace de Hölder. Pour cela on utilise la technique de la preuve du théorème de Kolmogorov basée sur les injections de Sobolev qu'on peut trouver dans [13] par exemple. Ensuite, pour passer à l'estimation de $a_N - a$, on utilise le théorème de Fernique, comme dans la preuve de la Proposition 1.2.2, mais en estimant les normes ; plus précisément on montre que a_N^{max} et $1/a_N^{min}$ sont bornés dans L^p. On en déduit donc une estimation de $a_N - a$, on notera que c'est une estimation uniforme en x pour ω fixé, contrairement à (1.8) avec le théorème de Mercer. Il en résulte donc à ω fixé une estimation de $u_N - u$ en norme $H_0^1(D)$, en utilisant une fois encore le caractère borné de a_N^{max} et $1/a_N^{min}$ dans L^p et un résultat classique de continuité de l'application qui au coefficient a muni de la norme uniforme associe la solution u munie de la norme $H_0^1(D)$. Le résultat de convergence presque-sûre est obtenu à l'aide du lemme de Borel-Cantelli.

Comme notre objectif général est de calculer la loi de u, on s'est intéressé à l'erreur faible commise en approchant u_N par u. On montre en particulier que l'ordre faible est le double de l'ordre fort, la loi de u_N converge plus vite vers la loi de u que les trajectoires de u_N vers les trajectoires de u. Pour cette estimation d'erreur faible, on se place dans le cas où la dimension spatiale d est $1, 2$ ou 3. Toujours sous l'hypothèse 1.2.4, on a l'estimation d'erreur faible suivante :

Théorème 1.2.8. *Pour toute fonction $\varphi \in \mathcal{C}^4(\mathbb{R}, \mathbb{R})$ dont les dérivées sont bornées, il existe une constante $C_{1.2.8}(\varphi, p)$ telle que pour tout $N \in \mathbb{N}$*

$$\|\mathbb{E}[\varphi(u_N)] - \mathbb{E}[\varphi(u)]\|_{L^p(D)} \leq C_{1.2.8}(\varphi, p) R_N^0, \tag{1.20}$$

pour $p \leq \infty$ si $d = 1$, $p < \infty$ si $d = 2$, et $p \leq \frac{3}{2}$ si $d = 3$.

On remarquera que l'erreur faible est de l'ordre de R_N^0, alors que l'erreur forte était presque de l'ordre de $\sqrt{R_N^0}$ (en supposant qu'on peut prendre α aussi proche de 0 que l'on veut, ce qui est le cas pour les exemples qui suivent), d'où le fait que l'ordre faible est le double de l'erreur forte. On vérifie numériquement que cette estimation de l'ordre faible est optimale (voir ci-dessous dans le cas d'une covariance exponentielle).

La preuve de ce résultat utilise des techniques sensiblement différentes de celles utilisées dans la preuve du résultat pour l'erreur forte. L'indépendance des variables aléatoires Y_n joue un rôle crucial dans le preuve de ce résultat d'erreur faible, alors que l'on a seulement utilisé le caractère décorrélé de celles-ci pour l'erreur forte. Pour obtenir l'estimation d'erreur faible, on remarque qu'on peut voir $\tilde{u}_N(y, x) - \tilde{u}(y, x)$ (avec les notations du paragraphe "Méthodes spectrales stochastiques" de la section 1.1.3) comme la différence entre des valeurs de $\tilde{u}$ correspondant à deux valeurs de y différentes, on va donc considérer un développement de Taylor (à l'ordre 2) de $\tilde{u}$ par rapport à y et estimer les dérivées de $\tilde{u}$ par rapport à y. Les variables aléatoires Y_n sont indépendantes, centrées et réduites, on obtient que les termes du premier ordre et ceux du second ordre correspondant à des dérivées croisées ($i \neq j$) sont nuls en moyenne, ce qui permet de gagner par rapport

à l'ordre fort. Plus précisément, formellement et dans le cas où φ est l'identité (pour simplifier l'exposé) on a le développement à l'ordre 2 suivant :

$$\begin{aligned} u(\omega,x)-u_N(\omega,x) &= \tilde{u}(Y_1(\omega),...,Y_N(\omega),Y_{N+1}(\omega),...,x)-\tilde{u}(Y_1(\omega),...,Y_N(\omega),0,...,x) \\ &= \sum_{i>N}\frac{\partial\tilde{u}}{\partial y_i}(Y_1(\omega),...,Y_N(\omega),0,...,x)Y_i(\omega) \\ &+ \frac{1}{2}\sum_{i,j>N}\frac{\partial^2\tilde{u}}{\partial y_i\partial y_j}(Y_1(\omega),...,Y_N(\omega),0,...,x)Y_i(\omega)Y_j(\omega)+... \end{aligned}$$

Les termes du premier ordre et ceux du second ordre avec $i\neq j$ sont d'espérance nulle et finalement, le terme dominant dans l'erreur est

$$\sum_{i>N}\mathbb{E}\left[\frac{\partial^2\tilde{u}}{\partial y_i^2}(Y_1(\omega),...,Y_N(\omega),0,...,x)\right].$$

Pour conclure on utilise l'estimation suivante des dérivées de u par rapport aux y_i : pour tout entier k, il existe une constante $C(k)$ telle que pour tout $N\in\mathbb{N}$, pour tout multi-indice $\alpha\in\mathbb{N}^N$ de longueur k, on ait

$$\left\|\frac{\partial^\alpha\tilde{u}_N}{\partial y^\alpha}\right\|_{H_0^1(D)}\leq C(k)\sqrt{\frac{\tilde{a}_N^{max}(y)}{\tilde{a}_N^{min}(y)}}||\tilde{u}_N(y)||_{H_0^1}\prod_{i\in\mathbb{N}}\sqrt{\lambda_i^{\alpha_i}}\|b_i\|_\infty^{\alpha_i}.$$

C'est le même type d'estimations que celles utilisées pour montrer l'analyticité de u par rapport à y dans [2, 1], ici il faut néanmoins noter qu'on ne peut pas avoir d'estimation uniforme par rapport à y, mais on utilise le fait que a_N^{max} et $1/a_N^{min}$ sont bornés dans L^p qu'on avait déjà utilisé pour la preuve de l'estimation d'erreur forte. On notera que pour passer de l'argument formel à un argument rigoureux, il faut estimer le reste intégral, ce qui n'est pas immédiat car on perd l'indépendance dans le reste intégral.

On remarque que dans l'argument brièvement décrit ci-dessus on utilise la norme H_0^1 et le résultat du Théorème 1.2.8 est vrai avec cette norme. Ceci est vrai car nous avons pris une fonctionnelle très simple : $\varphi(u(x))=u(x)$. Pour des fonctionnelles plus générales, on pourrait obtenir une majoration de l'erreur faible en norme $\mathcal{C}^1$ (et donc en norme H^1) mais la preuve nécessite des estimations très fines de u_N et de ses dérivées par rapport aux y_i en norme $C^{1,\beta}$.

Pour conclure, on donne deux cas importants dans lesquels ces estimations s'appliquent.

Proposition 1.2.9. *Les Hypothèses 1.2.4 et 1.2.6 sont vérifiées dans le cas d'une covariance exponentielle (définie par* (1.3)*) avec la norme* 1*, sur un domaine pavé, et dans le cas d'une fonction de covariance* cov *analytique sur* $\bar{D}\times\bar{D}$*.*

Tout d'abord l'Hypothèse 1.2.4 est vérifiée dans le cas d'une covariance exponentielle pour la norme 1. En effet, pour le cas de la dimension 1, on a vu que les couples propres sont définis quasi-explicitement par (1.9) et (1.10). On peut donc voir que les vecteurs propres sont continûment différentiables et vérifient $\|b_n\|_\infty\leq 2\sqrt{2}$ et $\|\nabla b_n\|_\infty\leq 4\sqrt{2}\pi n$ pour tout $n\in\mathbb{N}$. De plus

$$\lambda_n \underset{n\to+\infty}{\sim}\frac{2\sigma^2}{l\pi^2n^2}.$$

Donc l'Hypothèse (1.2.4) est vérifiée pour tout $\alpha\in]0,1/2[$. On a alors $R_n^\alpha=O(N^{\frac{2\alpha-1}{2}})$. On en déduit en particulier que l'Hypothèse 1.2.6 est vérifiée. Il est important de noter que la convergence des λ_n se détériore quand on diminue la longueur de corrélation ℓ. En effet on remarque sur la Figure 1.1 tout d'abord que $1/\ell$ est en facteur dans l'équivalent de λ_n, de plus on peut observer numériquement que la décroissance asymptotique des λ_n n'est atteinte qu'après un palier, dont la taille est empiriquement de l'ordre de $1/\ell$.

En dimension supérieure, on utlise le fait que les couples propres s'obtiennent en tensorisant les couples propres de la dimension 1.

Les Hypothèses 1.2.4 et 1.2.6 sont également vérifiées dans le cas d'une fonction de covariance analytique (par exemple le cas d'une covariance gaussienne, c'est-à-dire (1.4) avec $\delta=2$). Pour le montrer, on utlise le résultat suivant qui découle d'un résultat prouvé dans [24].

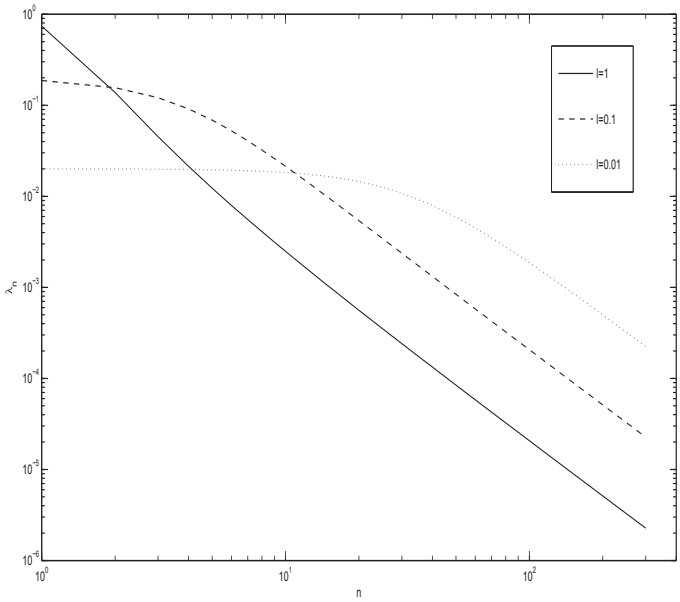


FIG. 1.1 – λ_n en fonction de n, en échelle logarithmique, pour $\sigma = 1$ et différentes valeurs de l .

Théorème 1.2.10. *Supposons la fonction de covariance* cov *analytique sur* $\bar{D} \times \bar{D}$*, alors les fonctions propres* b_n *sont* $\mathcal{C}^1$ *sur* $\bar{D}$*, et il existe des constantes* c_1, c_2 *telles que pour tout* $n \in \mathbb{N}$*,*

$$\lambda_n \leq c_1 e^{-c_2 n^{1/d}}.$$

Pour tout $s > 0$*, il existe une constante* c_s *telle que pour tout* $n \in \mathbb{N}$*,*

$$\|b_n\|_\infty \leq c_s |\lambda_n|^{-s} \quad et \quad \|\nabla b_n\|_\infty \leq c_s |\lambda_n|^{-s}.$$

Ce résultat implique alors que l'Hypothèse 1.2.4 est vérifiée pour tout $\alpha \in]0,1[$ et l'Hypothèse 1.2.6 est également vérifiée.

On peut donc appliquer les Théorèmes 1.2.7 et 1.2.8 dans ces deux cas. En particulier, dans le cas d'une covariance exponentielle définie par (1.3) pour la norme 1 sur un pavé on obtient une majoration pour l'erreur forte en $O(N^{\alpha-1/2})$ pour tout $0 < \alpha < 1/2$ et pour l'erreur faible en $O(N^{-1})$. On observe numériquement sur la Figure 1.2 une erreur faible en $1/N$, ce qui montre que notre estimation d'erreur faible est optimale. On obtient des estimations similaires en dimension 2.

1.2.3 Erreur éléments finis

Dans cette partie, on présente des estimations d'erreur éléments finis pour l'équation (1.18) qui seront utilisées dans les deux parties suivantes pour l'analyse numérique de la méthode de collocation stochastique et de la méthode de Monte-Carlo multi-niveaux qu'on définira par la suite. Dans cette partie, on reprend les preuves des estimations d'erreur éléments finis déterministes classiques (dans le cas très classique où le coefficient a est régulier, à savoir $\mathcal{C}^1$, et dans le cas où le coefficient a n'est que $\mathcal{C}^{0,\alpha}$) en rendant explicite la dépendance par rapport à ω. En utilisant une fois encore des estimations sur a basées sur le théorème de Fernique, on en déduit ensuite des estimations d'erreur éléments finis en normes $L^p(\Omega, H_0^1(D))$ et $L^p(\Omega, L^2(D))$,

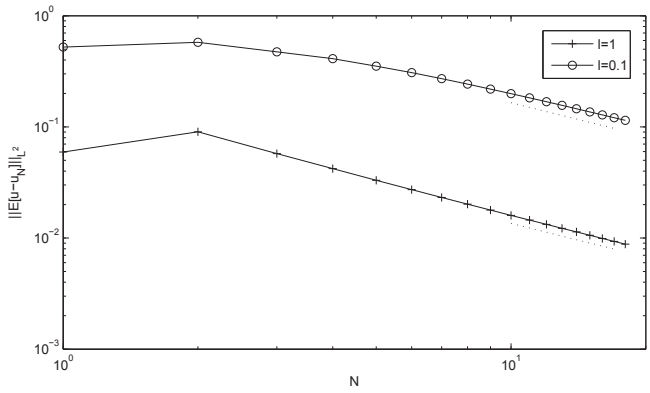


FIG. 1.2 – $\|\mathbb{E}[u-u_N]\|_{L^2}$ en fonction de N, en échelle logarithmique, pour $\sigma = 1$ dans les cas $l = 1$ et $l = 0,1$. Les pointillés indiquent une pente de -1.

que l'on complète avec l'estimation de l'erreur d'intégration numérique inhérente à la méthode d'éléments finis. De la même manière, on obtient des estimations éléments finis pour le champ tronqué a_N. On notera que dans ce cas, comme le champ approché a_N est plus régulier que a, on obtient a priori une estimation d'erreur éléments finis avec l'ordre maximal, mais une constante dépendant de N. On peut obtenir une constante indépendante de N si on a davantage de régularité sur les vecteurs propres et de décroissance sur les valeurs et vecteurs propres, ou si on se contente d'un ordre pour les éléments finis qui correspond à la régularité limite de a (dans le cas d'une covariance exponentielle définie par (1.3) par exemple).

Commençons par donner les résultats de régularité elliptique qui sont à la base des estimations d'erreur éléments finis, on a choisi de travailler avec des résultats de régularité dans des espaces de Sobolev, et on reprend donc les résulats classiques dans lesquels on détaille la dépendance par rapport au paramètre aléatoire ω qui est cruciale pour nous. On choisit un second membre dans L^2 pour simplifier cet exposé.

Théorème 1.2.11. *Si, pour presque tout ω, $a(\omega,\cdot) \in \mathcal{C}^1(\bar{D})$, alors il existe une constante $C_{2.2.1}$ telle que pour presque tout ω, $u(\omega,\cdot) \in H^2(D)$ avec*

$$\|u(\omega,\cdot)\|_{H^2(D)} \leq C_{2.2.1} \frac{a^{max}(\omega)\|a(\omega,\cdot)\|_{\mathcal{C}^1(\bar{D})}}{a^{min}(\omega)^3} \|f\|_{L^2(D)}. \tag{1.21}$$

Si, pour presque tout ω, $a(\omega,\cdot) \in \mathcal{C}^{0,t}(\bar{D})$ pour $0 < t < 1$, alors pour tout $0 < s < t$ avec $s \neq 1/2$, il existe une constante $C'_{2.2.1}$ telle que pour presque tout ω, $u(\omega,\cdot) \in H^{1+s}(D)$ avec

$$\|u(\omega,\cdot)\|_{H^{1+s}(D)} \leq C'_{2.2.1} \frac{a^{max}(\omega)\|a(\omega,\cdot)\|_{\mathcal{C}^{0,t}(\bar{D})}}{a^{min}(\omega)^3} \|f\|_{L^2(D)}. \tag{1.22}$$

Pour obtenir ces résultats, on a suivi les preuves classiques, qu'on peut trouver dans [9, 39], utilisant la méthode des translations, preuves en trois étapes où on commence par les cas de l'espace entier et du demi espace, puis on s'y ramène à l'aide de cartes locales.

A partir de ces résultats, on va pouvoir obtenir des estimations d'erreur pour la méthode des éléments finis appliquée à l'équation (1.18) pour presque tout ω. Le domaine D étant supposé $\mathcal{C}^2$ (pour obtenir plus facilement les résultats de régularité), on l'approche par un domaine polygonal D_h dont la distance à D est inférieure à h, pour simplifier on suppose également D convexe. On considère alors comme précédemment un espace standard d'éléments finis V_h sur D_h de fonctions continues, linéaires par morceaux, associé à une famille régulière de triangulations de D dont on note h le diamètre maximum du maillage. Pour presque tout ω, on définit $u_h(\omega,\cdot)$ l'approximation par éléments finis de $u(\omega,\cdot)$ dans V_h. On a alors les estimations d'erreur suivantes, les deux premières dans le cadre des hypothèses faites dans le paragraphe 1.2.1, les deux suivantes avec des hypothèses de régularité supplémentaires.

Théorème 1.2.12. *Soit $s < 1/2$, pour tout $p > 0$, il existe des constantes $C_{5.4.3}(p,s)$ et $C'_{5.4.3}(p,s)$ telles que*

$$\|u - u_h\|_{L^p(\Omega, H_0^1(D))} \leq C_{5.4.3}(p,s)\|f\|_{L^2(D)}h^s \quad et \quad \|u - u_h\|_{L^p(\Omega, L^2(D))} \leq C'_{5.4.3}\|f\|_{L^2(D)}h^{2s}. \tag{1.23}$$

Cette estimation reste vraie pour $1/2 < s < 1$ si on fait l'hypothèse supplémentaire qu'il existe $t > s$ tel que $a \in L^q(\Omega, \mathcal{C}^{0,t}(\bar{D}))$ pour tout $q > 0$.
Si on fait l'hypothèse supplémentaire que $a \in L^q(\Omega, \mathcal{C}^1(\bar{D}))$ pour tout $q > 0$, alors pour tout $p > 0$, il existe des constantes $D_{5.4.3}(s)$ et $D'_{5.4.3}(s)$ telles que

$$\|u - u_h\|_{L^p(\Omega, H_0^1(D))} \leq D_{5.4.3}(s)\|f\|_{L^2(D)}h \quad et \quad \|u - u_h\|_{L^p(\Omega, L^2(D))} \leq D'_{5.4.3}(s)\|f\|_{L^2(D)}h^2. \tag{1.24}$$

On notera que dans le cas d'une covariance exponentielle (1.3), on obtient seulement un ordre inférieur à $1/2$ en norme H_0^1 et inférieur à 1 en norme L^2. On peut voir que cette estimation d'erreur forte est optimale sur la Figure 1.3 qui correspond à l'erreur éléments finis trajectorielle en norme L^2 dans le cas d'une covariance exponentielle.

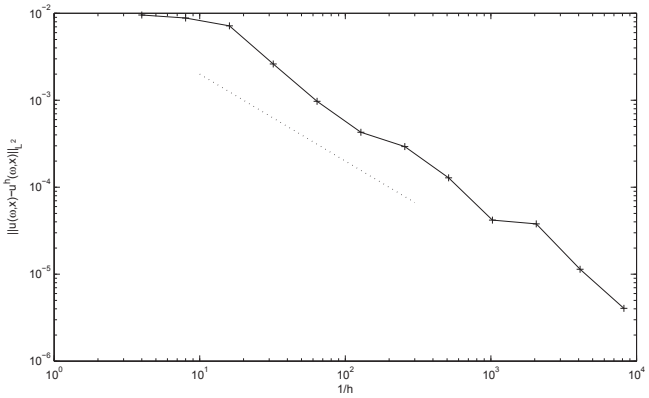


FIG. 1.3 – $\|u(\omega,x) - u^h(\omega,x)\|_{L^2(0,1)}$ en fonction de h, en échelle logarithmique. Les pointillés indiquent une pente de -1.

Tout d'abord, pour montrer la première estimation, on note que les hypothèses faites dans le paragraphe 1.2.1 impliquent que pour tout $s < 1/2$, $a \in L^q(\Omega, \mathcal{C}^{0,s}(\bar{D}))$ pour tout $q > 0$, la preuve de ce résultat utilise les techniques vues pour les Propositions 1.2.1 et 1.2.2. Pour montrer les estimations, on utilise les résultats de régularité du théorème 1.2.11, ainsi que les techniques usuelles pour les estimations d'erreur éléments finis : lemme de Céa, résultat d'approximation de fonctions H^{1+s} dans V_h, approximation de D par D_h et méthode de dualité pour l'estimation L^2. On conclut en utilisant que $a \in L^q(\Omega, \mathcal{C}^{0,t}(\bar{D}))$ et $1/a^{min} \in L^q(\Omega)$ pour tout $q > 0$ et $t > s$.

On complète maintenant cette estimation d'erreur éléments finis par une estimation prenant en compte l'erreur de quadrature. On note $v_h(\omega, \cdot)$ la solution approchée obtenue en utilisant une méthode du point milieu comme méthode d'intégration numérique pour calculer les intégrales apparaissant dans la méthode d'éléments finis, c'est-à-dire que $v_h(\omega, \cdot)$ est la fonction de V_h vérifiant : pour tout $w \in V_h$

$$\sum_{\tau \in \mathcal{T}_h} a(\omega, x_\tau) \int_\tau \nabla w(x) \nabla v_h(\omega, x) dx = \int_{D_h} f(x) v_h(x) dx, \tag{1.25}$$

où on a noté x_τ le centre du simplexe $\tau \in \mathcal{T}_h$, et où $\mathcal{T}_h$ est la triangulation régulière de D_h associée à V_h. On a alors l'estimation d'erreur suivante :

Corollaire 1.2.13. *Pour $0 < s < 1/2$, et $p > 0$, il existe une constante $C_{1.2.13}(s,p)$ telle que*

$$\|u - v_h\|_{L^p(\Omega, H_0^1(D))} \leq C_{1.2.13} \|f\|_{L^2(D)} h^s. \tag{1.26}$$

On notera que, comme dans le cas du Théorème 1.2.12, on peut améliorer cette estimation si on suppose davantage de régularité sur a. Pour montrer cette estimation, on utilise de manière classique le premier lemme de Strang, en explicitant la dépendance par rapport à ω, puis on conclut en utilisant que $a \in L^q(\Omega, \mathcal{C}^{0,t}(\bar{D}))$ et $1/a^{min} \in L^q(\Omega)$ pour tout $q > 0$ et $t > s$.

On s'intéresse maintenant aux estimations d'erreur éléments finis pour le champ tronqué a_N. On peut bien entendu appliquer les estimations ci-dessus (à savoir (1.3),(1.5),(1.26)) au champ approché a_N, on obtient alors des constantes dépendant de N, qui en général n'ont aucune raison d'être des fonctions bornées de N. Dans le cas d'une covariance exponentielle (définie par (1.3)), cette constante augmente avec N comme on peut le voir sur la Figure 1.4.

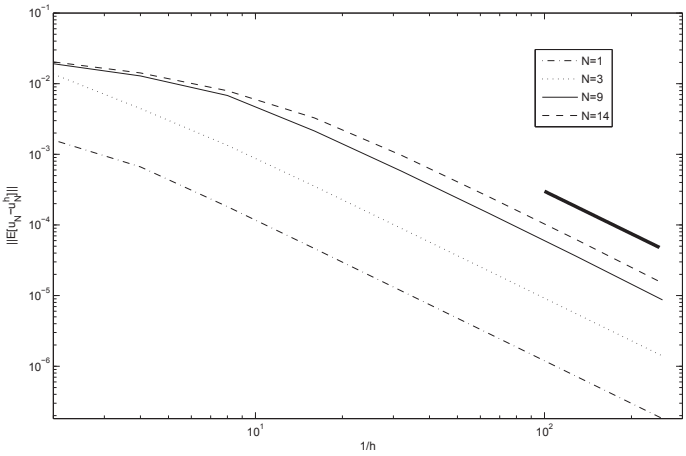


FIG. 1.4 – $\|\mathbb{E}[u_N^h] - \mathbb{E}[u_N]\|_{L^2}$ en fonction de h, pour $N = 1, 3, 9, 14$. Le trait épais indique une pente de -2.

On peut obtenir une estimation d'erreur éléments finis indépendante de N avec un ordre plus faible, correspondant à la régularité de la limite u.

Proposition 1.2.14. *On suppose l'Hypothèse 1.2.4 vérifiée. Soit $s < 1/2$, pour tout $p > 0$, il existe des constantes $C_{1.2.14}(p,s)$ et $C'_{1.2.14}(p,s)$ telles que pour tout $N \in \mathbb{N}$*

$$\|u_N - u_N^h\|_{L^p(\Omega, H_0^1(D))} \leq C_{1.2.14}(p,s) \|f\|_{L^2(D)} h^s \quad \text{et} \quad \|u_N - u_N^h\|_{L^p(\Omega, L^2(D))} \leq C'_{1.2.14} \|f\|_{L^2(D)} h^{2s}. \tag{1.27}$$

On notera que ce résulat s'applique en particulier dans les cas d'une covariance exponentielle et d'une covariance analytique, d'après la Proposition 1.2.9. Dans le cas d'une covariance exponentielle, on peut voir sur la Figure 1.5 qu'en fait la courbe de convergence en norme L^2 fait apparaître deux zones, pour des valeurs de h supérieures à $1/N$ on a presque un ordre 1, conformément à (1.23) et correspondant à l'ordre limite quand N tend vers l'infini, et pour des valeurs de h inférieures à $1/N$ on a l'ordre asymptotique de 2, correspondant à l'ordre maximal car u_N est régulière.

Néanmoins, pour un covariance analytique, on peut obtenir une meilleure estimation. En effet on peut également obtenir une estimation d'erreur indépendante de N avec l'ordre maximal en supposant davantage de régularité et de décroissance sur les couples propres (λ_n, b_n) (ce qui implique davantage de régularité sur la limite u).

Proposition 1.2.15. *On suppose que les vecteurs propres b_n appatiennent à $\mathcal{C}^2(\bar{D})$, que la série*

$$\sum_{n \geq 0} \lambda_n \|\nabla b_n\|_\infty^2$$

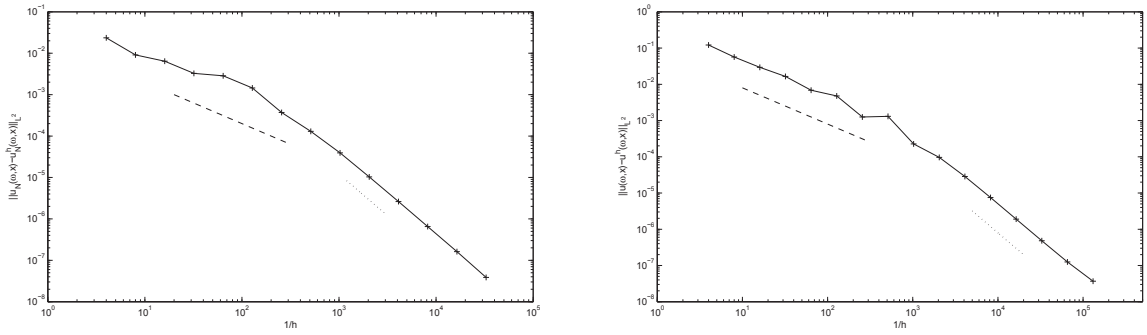


FIG. 1.5 – $\|u_N(\omega,x) - u_N^h(\omega,x)\|_{L^2(0,1)}$ en fonction de h, en échelle logarithmique, pour $N = 500$ (à gauche) et $N = 2000$ (à droite). Les pointillés indiquent une pente de -2, les pointillés larges indiquent une pente de -1.

est convergente, et qu'il existe $0 < \theta < 1$ *tel que la série*

$$\sum_{n\geq 0} \lambda_n \|\nabla b_n\|_\infty^{2(1-\theta)} \|D^2 b_n\|_\infty^{2\theta}$$

soit convergente. Alors pour tout $p > 0$*, il existe une constante* $C_{1.2.15}(p)$ *telle que*

$$\|u_N - u_N^h\|_{L^p(\Omega, H_0^1(D))} \leq C_{1.2.15}(p) \|f\|_{L^2(D)} h. \tag{1.28}$$

Pour montrer ce résultat, on utilise le résultat de régularité (1.24) appliqué au champ tronqué a_N, et on montre que l'hypothèse supplémentaire implique que a_N est borné dans $L^p(\Omega, \mathcal{C}^1(D))$ et $1/a_N^{min}$ est borné dans $L^p(\Omega)$ pour tout $p > 0$, en utlisant des techniques similaires à celles vues précédemment. Ce résultat s'applique en particulier dans le cas d'une covariance analytique.

Nous allons maintenant nous intéresser à l'analyse numérique des méthodes de collocation stochastiques et de Monte-Carlo multi-niveaux, dans laquelle les estimations d'erreur qu'on vient de voir jouent un rôle important.

1.2.4 Méthode de collocation stochastique

Dans ce paragraphe, on s'intéresse à la méthode de collocation stochastique telle qu'elle est présentée dans [1] et que nous l'avons présentée rapidement dans le paragraphe "Méthodes spectrales stochastiques" de la section 1.1.3. Nous allons étendre l'estimation d'erreur vue dans le Théorème 1.1.4 et prouvée dans [1], obtenue sous l'Hypothèse 1.1.1, au cas d'un champ lognormal peu régulier, c'est-à-dire sous les hypothèses de cette partie, données dans la section 1.2.1. L'erreur peut être découpée en trois termes, un terme correspondant à l'erreur de troncature, un terme correspondant à l'erreur de collocation et un terme correspondant à l'erreur d'éléments finis. L'estimation de l'erreur de collocation à proprement parler doit alors être adaptée à cause du caractère non uniformément borné et coercif de a par rapport au paramètre aléatoire ω. L'estimation d'erreur éléments finis doit être adpatée à cause de ce caractère non uniformément borné et coercif et à cause du manque de régularité, elle est donnée directement par les résultats du paragraphe 1.2.3. L'erreur de troncature est donnée directement par les résultats du paragraphe 1.2.2.

Théorème 1.2.16. *Pour tout* $0 < s < 1/2$*,* $N \in \mathbb{N}$*, il existe une constante* $C_{1.2.16}(s)$*, des constantes positives* $r_1, ..., r_N$*, une constante* $C'_{1.2.16}(N)$ *et une constante* $C''_{1.2.16}$ *telles que*

$$\|\mathbb{E}[u - u_N^{h,p}]\|_{H^1(D)} \leq C_{1.2.16}(s) h^s \|f\|_{L^2(D)} + C'_{1.2.16}(N) \sum_{n=1}^{N} \sqrt{p_n} e^{-r_n \sqrt{p_n}} + C''_{1.2.16} R_N^0. \tag{1.29}$$

Pour prouver cette estimation, on découpe l'erreur en trois termes comme évoqué ci-dessus.

$$\|\mathbb{E}[u - u_N^{h,p}]\|_{H^1(D)} \leq \|\mathbb{E}[u - u_N]\|_{H^1(D)} + \|\mathbb{E}[u_N - u_N^h]\|_{H^1(D)} + \|\mathbb{E}[u_N^h - u_N^{h,p}]\|_{H^1(D)}.$$

Le premier terme, qui est une erreur de troncature, est majoré en utilisant le Théorème 1.2.8. Le second terme, qui est une erreur d'éléments finis pour le champ tronqué, est majoré en utilisant la Proposition 1.2.14. Le dernier terme est une erreur d'interpolation, pour le majorer on adapte légèrement la preuve de [1] pour tenir compte du caractère non-uniformément borné et coercif. Voici les grandes étapes de la preuve. On commence par montrer que pour chaque variable stochastique mono-dimensionnelle y_n, l'application $y_n \mapsto \tilde{u}_N(y_n, y_n^*, \cdot)$ à valeurs dans l'espace des fonctions continues sur $\mathbb{R}^{N-1}$ à valeurs dans $H_0^1(D)$, muni d'un poids σ_n sur $\mathbb{R}^{N-1}$, admet un prolongement analytique sur toute bande autour de l'axe réel, à distance strictement inférieure à $\frac{1}{2\sqrt{\lambda_n}\|b_n\|_\infty}$ de cet axe. On obtient en même temps une estimation de la norme de $\tilde{u}_N$ dans cet espace de fonctions continues sur $\mathbb{R}_{N-1}$ avec le poids σ_n. Pour obtenir ce résultat, on estime les dérivées de $\tilde{u}_N$ par rapport à y_n. On décompose ensuite l'erreur d'interpolation pour se ramener à une interpolation mono-dimensionnelle, l'erreur d'interpolation mono-dimensionnelle est alors majorée par une erreur d'approximation, qu'on estime grâce à un résultat de [41] en utilisant le résultat d'analyticité évoqué précédemment.

1.2.5 Méthode de Monte-Carlo multi-niveaux

Dans ce paragraphe, on utilise les résultats d'estimation d'erreur éléments finis du paragraphe 1.2.3 pour obtenir l'analyse numérique de la méthode de Monte-Carlo multi-niveaux telle qu'elle a été proposée dans [12]. Une méthode de Monte-Carlo multi-niveaux similaire a été proposée dans [4], où l'analyse numérique de la méthode est proposée sous une hypothèse similaire à l'Hypothèse 1.1.1. Dans [12], on trouve des résultats numériques 2d qui illustrent l'efficacité de la méthode dans le cas d'un champ lognormal avec covariance exponentielle (1.3) vu dans le paragraphe "Ecoulement" de la section 1.1.2.

Commençons par présenter brièvement la méthode. Notant encore u la solution de (1.18), on souhaite calculer l'espérance d'une fonction de u à valeurs réelles, $Q = \mathcal{G}(u)$. Comme dans le paragraphe 1.2.3, on note $u_h(\omega, \cdot)$ l'approximation par éléments finis de $u(\omega, \cdot)$ dans l'espace d'éléments finis V_h associé à la grille $\mathcal{T}_h$ (avec les même hypothèses et notations que dans le paragraphe 1.2.3). On peut alors approcher Q par $Q_h = \mathcal{G}(u_h)$. Pour une estimation $\hat{Q}_h$ de $\mathbb{E}[Q_h]$, on note $e(\hat{Q}_h) = (\mathbb{E}[(\hat{Q}_h - \mathbb{E}[Q])^2])^{1/2}$ l'erreur associée en norme L^2. Le ε-coût de calcul $\mathcal{C}_\varepsilon(\hat{Q}_h)$ associé est alors défini comme le nombre d'opérations élémentaires nécessaire pour obtenir une erreur L^2, $e(\hat{Q}_h) \leq \varepsilon$. Avec ces notations, la méthode de Monte-Carlo classique, telle qu'on l'a présentée dans le paragraphe "Méthodes de type Monte-Carlo" de la section 1.1.3 correspond à

$$\hat{Q}_{h,N}^{MC} = \frac{1}{N} \sum_{i=1}^{N} Q_h^i,$$

où les $(a^i)_{1 \leq i \leq N}$ sont des réalisations indépendantes de a et les $(Q_h^i)_{1 \leq i \leq N}$ les approximations éléments finis de la solution associée. L'idée de base de la méthode de Monte-Carlo multi-niveaux est d'utiliser plusieurs grilles $(\mathcal{T}_{h_l})_{l=0,\ldots,L}$, les calculs sur les grilles grossières permettant de réduire grandement la variance. Plus précisément, notant $h_0 \geq h_1 \geq \ldots \geq h_L$ les différents pas de maillage utilisés et $\mathcal{T}_h = \mathcal{T}_{h_L}$ la grille la plus fine, on commence par faire la remarque élémentaire suivante :

$$\mathbb{E}[Q_h] = \mathbb{E}[Q_{h_0}] + \sum_{l=1}^{L} \mathbb{E}[Q_{h_l} - Q_{h_{l-1}}].$$

Le principe de la méthode de Monte-Carlo multi-niveaux est maintenant d'approcher de manière indépendante ces espérances à l'aide d'une méthode de Monte-Carlo standard, tirant profit du fait que plus on utlise des grilles fines, plus la variance de la différence correspondante est réduite ; donc la convergence de la méthode de Monte-Carlo est accélérée et les calculs sur les grilles grossières sont peu coûteux. Plus précisément, notant $Y_l = Q_{h_l} - Q_{h_{l-1}}$ pour $1 \leq l \leq L$ et $Y_0 = Q_{h_0}$, on définit l'approximation

$$\hat{Q}_{h,N_l}^{ML} = \sum_{l=0}^{L} \hat{Y}_{l,n_l}^{MC} = \sum_{l=0}^{L} \frac{1}{N_l} \sum_{i=1}^{N_l} Y_l^i,$$

où les $Y_l^i = Q_{h_l}^i - Q_{h_{l-1}}^i$ sont calculés à partir d'une même réalisation $a^{i,l}$ de a pour les deux grilles, les $(a^{i,l})_{0\le l\le L, 1\le i\le N_l}$ étant des réalisations indépendantes de a. On peut dès à présent remarquer que l'erreur $e(\hat{Q}^{ML}_{h,\{N_l\}})$ se décompose en deux termes, un terme contenant les erreurs des Monte-Carlo, et un terme d'erreur de discrétisation spatiale correspondant à la grille la plus fine :

$$e(\hat{Q}^{ML}_{h,\{N_l\}})^2 = \sum_{l=0}^{L} \frac{\mathbb{V}[Y_l]}{N_l} + (\mathbb{E}[Q - Q_h])^2.$$

On note par ailleurs $\mathcal{C}_l$ le coût pour obtenir une réalisation de Q_{h_l}.

Nous allons maintenant établir l'analyse numérique de cette méthode, une fois encore dans un cadre similaire à celui présenté dans le paragraphe "Ecoulement" de la section 1.1.2. Pour commencer, on rappelle un théorème général pour la méthode de Monte-Carlo multi-niveaux, dont on peut trouver la preuve dans [12, 35], et qui permet de majorer le ε-coût de la méthode.

Théorème 1.2.17. *On suppose qu'il existe des constantes $\alpha, \beta, \gamma > 0$ telles que $\alpha \ge \frac{1}{2}\min(\beta, \gamma)$ et*

i) $|\mathbb{E}[Q_h - Q]| = O(h^\alpha)$,

ii) $\mathbb{V}[Q_{h_l} - Q_{h_{l-1}}] = O(h_l^\beta)$,

iii) $\mathcal{C}_l = O(h_l^{-\gamma})$,

alors, pour tout $\varepsilon < e^{-1}$, il existe une valeur L et une suite $\{N_l\}_{0\le l\le L}$ telle que $e(\hat{Q}^{ML}_{h,\{N_l\}}) < \varepsilon$ et

$$\mathcal{C}_\varepsilon(\widehat{Q}^{\mathrm{MLMC}}_{h,\{N_\ell\}}) \;=\; \begin{cases} O(\varepsilon^{-2}), & \text{si } \beta > \gamma, \\ O(\varepsilon^{-2}(\log\varepsilon)^2), & \text{si } \beta = \gamma, \\ O(\varepsilon^{-2-(\gamma-\beta)/\alpha}), & \text{si } \beta < \gamma. \end{cases}$$

Pour la méthode de Monte-Carlo classique, on a $\mathcal{C}_\varepsilon(\hat{Q}_h^{MC}) = O(\varepsilon^{-2-\gamma/\alpha})$.

On notera que le nombre de réalisations N_l au niveau l qui permet d'obtenir un coût inférieur à ε peut être approché numériquement, à partir de $\mathcal{C}_l$ et de la variance empirique de Y_l. Le nombre total de grilles L peut quant à lui être approché grâce à l'espérance empirique de Y_l.

Nous allons maintenant utiliser les estimations d'erreur éléments finis obtenues dans le paragraphe 1.2.3 pour majorer le ε-coût dans les cas où on s'intéresse aux quantités $Q = \mathcal{G}(u) = \|u\|^q_{H^1_0(D)}$ et $Q = \mathcal{G}(u) = \|u\|^q_{L^2(D)}$ pour $q \ge 1$, ceci dans le cadre des hypothèses du paragraphe 1.2.1.

Proposition 1.2.18. *1. Dans le cas où $Q = \mathcal{G}(u) = \|u\|^q_{H^1_0(D)}$ pour un $q \ge 1$, les hypothèses du Théorème 1.2.17 sont vérifiées pour $\alpha < 1/2$ et $\beta < 1$. Si on suppose de plus que $a \in L^p(\Omega, \mathcal{C}^{0,t}(\bar{D}))$ pour tout $p > 0$ et $1/2 \le t < 1$, alors on peut prendre $\alpha < t$ et $\beta < 2t$, et si on suppose $a \in L^p(\Omega, \mathcal{C}^1(\bar{D}))$ pour tout $p > 0$, alors on peut prendre $\alpha = 1$ et $\beta = 2$.*

2. Dans le cas où $Q = \mathcal{G}(u) = \|u\|^q_{H^1_0(D)}$ pour un $q \ge 1$, les hypothèses du Théorème 1.2.17 sont vérifiées pour $\alpha < 1$ et $\beta < 2$. Si on suppose de plus que $a \in L^p(\Omega, \mathcal{C}^{0,t}(\bar{D}))$ pour tout $p > 0$ et $1/2 \le t < 1$, alors on peut prendre $\alpha < 2t$ et $\beta < 4t$, et si on suppose $a \in L^p(\Omega, \mathcal{C}^1(\bar{D}))$ pour tout $p > 0$, alors on peut prendre $\alpha = 2$ et $\beta = 4$.

On notera, que dans les deux cas, on peut prendre des plus grandes valeurs de α et β en supposant davantage de régularité sur a.

On déduit entre autres de cette proposition les résultats présentés dans le tableau 1.6, en supposant que le coût de calcul de résolution du système linéaire associé à la méthode d'éléments finis est optimal, c'est-à-dire vérifie $\mathcal{C}_l \le h_l^{-d}\log(h_l^{-1})$.

On remarquera en particulier au vu du Tableau 1.6 que dans le cas de la covariance exponentielle en dimension $d > 1$, l'ordre de l'ε-coût de la méthode de Monte-Carlo multi-niveaux est le même que celui de la résolution d'un problème déterministe (avec même régularité).

La Figure 1.7 montre que la méthode de Monte-Carlo multi-niveaux est nettement plus efficace que la méthode de Monte-Carlo standard.

	$\|u\|_{H^1(D)}$		$\|u\|_{L^2(D)}$	
d	MC	MLMC	MC	MLMC
1	ε^{-3}	ε^{-2}	$\varepsilon^{-5/2}$	ε^{-2}
2	ε^{-4}	ε^{-2}	ε^{-3}	ε^{-2}
3	ε^{-5}	ε^{-3}	$\varepsilon^{-7/2}$	ε^{-2}

	$\|u\|_{H^1(D)}$		$\|u\|_{L^2(D)}$	
d	MC	MLMC	MC	MLMC
1	ε^{-4}	ε^{-2}	ε^{-3}	ε^{-2}
2	ε^{-6}	ε^{-4}	ε^{-4}	ε^{-2}
3	ε^{-8}	ε^{-6}	ε^{-5}	ε^{-3}

FIG. 1.6 – Majorations théoriques pour le ε-coût des méthodes de Monte-Carlo classique et multi-niveaux, d'après la Proposition 1.2.18 dans le cas d'un champ lognormal avec covariance gaussienne définie par (1.4) avec $\delta = 2$ (à gauche) et avec covariance exponentielle définie par (1.3) (à droite). (Pour simplifier, on note dans ce tableau ϵ^{-p}, au lieu de $\epsilon^{-p-\delta}$ pour tout $\delta > 0$.)

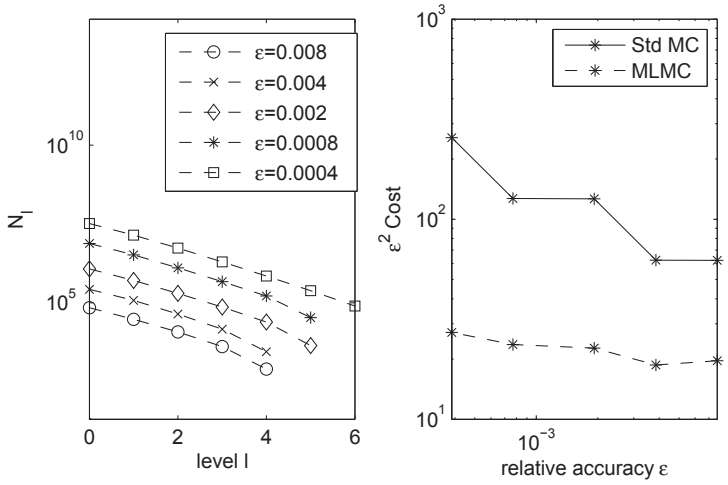


FIG. 1.7 – A gauche : Nombre d'échantillons N_ℓ par niveau. A droite : Coût renormalisé par ε^{-2} des approximations par les méthodes MLMC et MC , avec $\lambda = 0.1$ et $\sigma^2 = 3$. La grille la plus grossière est $h_0 = 1/16$.

1.3 Analyse numérique du couplage de l'équation d'écoulement et de l'équation d'advection-diffusion

Dans cette partie, on s'intéresse au couplage de l'équation d'écoulement (1.1) avec l'équation d'advection-diffusion (5.5) présentée dans la section 1.1.2. On rappelle qu'on souhaite calculer l'extension moyenne, et sa dérivée la dispersion moyenne, qui représente la vitesse moyenne à laquelle le panache de soluté s'étend. Dans un premier temps, on présente la méthode proposée dans [15], et dans un deuxième temps on en fait l'analyse numérique. Les résultats de cette partie sont détaillés dans le Chapitre 5.

1.3.1 Une méthode de Monte-Carlo particulaire probabiliste

Comme on s'intéresse au cas $\sigma \geq 1$, c'est-à-dire au cas où les incertitudes sont importantes, les méthodes de type petites perturbations ne semblent pas adaptées. Par ailleurs on s'intéresse particulièrement au cas $\ell \simeq 1$, ce qui implique que les méthodes spectrales stochastiques ne semblent pas non plus appropriées ; en effet dans ce cas la convergence du développement de Karhunen-Loève est lente, et commence par un palier important. Une méthode de type Monte-Carlo a donc été proposée pour ce couplage d'équations.

On considère donc N réalisations indépendantes du champ de perméabilité a : $(a^i)_{1\leq i\leq N}$. Pour chacune de ces réalisations, on calcule une approximation de l'extension S^i et de la dispersion $\mathcal{D}^i$ correspondantes, et à la fin les valeurs moyennes de S et $\mathcal{D}$ sont approchées par les moyennes empiriques des S^i et des $\mathcal{D}^i$. On

est donc ramené à calculer la solution du couplage de l'équation d'écoulement et de l'équation d'advection-diffusion dans le cas déterministe. On utilise des éléments finis pour calculer la pression hydraulique u, et on en déduit donc une approximation $\tilde{v}$ de la vitesse de Darcy $v = -a\nabla u$. Ensuite, puisqu'on est intéressé par des quantités statistiques (spatiales) et non par la concentration en tout point, et puisque le nombre de Péclet est supposé grand (ce qui signifie que l'advection domine) on choisit une méthode particulaire probabiliste, qui évite les phénomènes de diffusion numérique. Pour simplifier, on suppose que l'équation d'advection-diffusion est posée sur $\mathbb{R}^d$ tout entier. Rappelant que $\mathrm{div}(v) = 0$, l'EDP (1.5) est en fait une équation de Fokker-Planck associée à l'équation différentielle stochastique

$$dX(t) = v(X(t))dt + \sqrt{2D}dW(t),$$

ce qui signifie qu'en ajoutant à l'EDS et à l'EDP les conditions initiales correspondantes, la variable aléatoire $X(t)$ admet $c(x,t)$ pour densité par rapport à la mesure de Lebesgue. Par conséquent, le centre de masse s'exprime comme $G(t) = \mathbb{E}[X(t)]$ et l'extension comme $S(t) = \mathbb{E}[(X(t)-G(t))(X(t)-G(t))^t]$. On utilise donc finalement une méthode de Monte-Carlo et un schéma d'Euler pour l'EDS afin de calculer S, c'est-à-dire qu'on considère M réalisations indépendantes $(X_n^j)_{1\leq j\leq M}$ du schéma d'Euler de pas $\Delta t = T/n$ associé à l'EDS et on approche G par la moyenne empirique $\bar{G}$ des X_n^j et S par la moyenne empirique des $(X_n^j - \bar{G})(X_n^j - \bar{G})^t$. La dispersion $\mathcal{D}(t)$, qui est définie comme la dérivée de l'extension S est quant à elle approchée par le taux d'accroissement de S entre t et $t + \Delta s$.

1.3.2 Analyse numérique de la méthode

Malheureusement, nous n'avons pas été en mesure de faire l'analyse numérique de cette méthode dans le cadre des hypothèses du paragraphe 1.1.2. Tout d'abord, l'analyse numérique de cette méthode requiert que la solution de l'équation (1.1) soit suffisament régulière par rapport à la variable d'espace x. Avec les hypothèses de la partie 1.1.2, les trajectoires de a ne sont a priori que $\mathcal{C}^{0,\alpha}$ pour $\alpha < 1/2$, ce qui donne la même régularité pour les trajectoires de la vitesse v. Notre analyse requiert une régularité au moins $\mathcal{C}^{1,\beta}$ pour un $\beta > 0$, ce qui exclut le cas d'une covariance exponentielle, traitant par ailleurs les cas où la covariance est définie par (1.4) avec $\delta > 2 + 2\beta$ ou $\delta = 2$. Par ailleurs, notons que l'équation d'advection-diffusion est posée sur $\mathbb{R}^d$ tout entier et ce n'est a priori pas évident d'y étendre la vitesse qui n'est a priori définie que sur D, on va donc remplacer les conditions aux bords pour l'équation d'écoulement (1.1) par des conditions périodiques. Ceci ne modifie pas beaucoup le résultat final en pratique, car le domaine D sur lequel on calcule la pression est choisi suffisament grand pour que quasiment aucune particule n'en sorte. Enfin, comme c'est souvent le cas dans les travaux périodiques, on va supposer que le champ de perméabilité a est uniformément borné et coercif. Comme on l'a déjà vu ce n'est clairement pas le cas pour un champ lognormal, mais on peut imaginer pouvoir obtenir des résultats similaires sans cette hypothèse en utilisant les techniques vues précédemment. Pour conclure ces considérations nous amènent à l'hypothèse suivante utilisée pour les résultats ci-dessous.

Hypothèse 1.3.1. *On suppose que $a \in L^\infty(\Omega, \mathcal{C}_b^{1,\alpha}(\mathbb{R}^d))$ pour un certain $\alpha > 0$, que pour presque tout ω, $a(\omega,\cdot)$ est O-périodique et vérifie pour tout $x \in \mathbb{R}^d$,*

$$0 < a^{min} \leq a(\omega, x) \leq a^{max} < +\infty.$$

On considère alors l'équation d'écoulement modifiée suivante :

$$\left\{ \begin{array}{rcll} -\mathrm{div}(a(\omega,x)\nabla u(\omega,x)) & = & f(x), & \text{sur } \Omega \times \mathbb{R}^d, \\ \int_O u(\omega,x)dx & = & 0 & \text{sur } \Omega, \end{array} \right. \tag{1.30}$$

et u est presque sûrement O-périodique.

On a alors le résultat suivant.

Proposition 1.3.2. *L'équation (1.30) admet une unique solution $u \in L^\infty(\Omega, \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d))$.*

La régularité de la solution est obtenue en appliquant un résultat classique de régularité Hölderienne pour les EDP elliptiques, en vérifiant qu'on a une majoration uniforme par rapport à ω. On considère maintenant l'approximation u_h par éléments finis de u dans un espace d'éléments finis V_h, et on note encore $v = -a\nabla u$ et, de manière naturelle, $v_h = -a\nabla u_h$. On a alors $v \in L^\infty(\Omega, \mathcal{C}_b^{1,\alpha}(\mathbb{R}^d))$ et $v_h \in L^\infty(\Omega \times \mathbb{R}^d)$. On fait dorénavant l'hypothèse suivante sur l'espace d'éléments finis V_h.

Hypothèse 1.3.3. *On suppose qu'il existe une constante C telle que*

$$\|v - v_h\|_{L^\infty(\Omega,\mathbb{R}^d)} \leq Ch|\log(h)|.$$

On redéfinit maintenant l'équation d'avection-diffusion.

$$\left\{ \begin{array}{rcll} \frac{\partial c}{\partial t}(\omega,x,t) + v(\omega,x).\nabla c(\omega,x,t) - D\Delta c(\omega,x,t) & = & 0, & x \in \mathbb{R}^d \text{ and } t \in [O,T] \\ c(\omega,x,0) & = & c_0(x), & x \in \mathbb{R}^d, \end{array} \right. \tag{1.31}$$

où $c_0(x) = \mathbb{1}_R(x)$, comme défini dans le paragraphe Advection-diffusion de polluants de la section 1.1.2 et v est défini ci-dessus. Notant $(\Omega', \mathcal{F}', \mathbb{P}')$ un nouvel espace de probabilité dont la variable générique est notée ξ, on définit la solution de l'EDS associée à l'EDP (1.31) par

$$\left\{ \begin{array}{rcl} dX(\omega,\xi,t) & = & v(\omega, X(\omega,\xi,t))dt + \sqrt{2D}dW(\xi,t), \\ X(\omega,\xi,0) & = & X_0(\xi), \end{array} \right. \tag{1.32}$$

où W est un mouvement brownien d-dimensionnel et où X_0 admet c_0 pour densité. On a donc le résultat classique suivant :

Proposition 1.3.4. *Pour presque tout ω, l'équation* (1.31) *admet une unique solution $c(\omega,\cdot) \in \mathcal{C}(]0,T], \mathcal{C}^2(\mathbb{R}^d))$ et l'unique solution X de* (1.32) *est telle que pour tout t, $X(\omega,\cdot,t)$ admet $c(\omega,x,t)$ pour densité par rapport à la mesure de Lebesgue.*

On rappelle que l'extension moyenne s'exprime alors comme

$$S(t) = \mathbb{E}_\omega[\mathbb{E}_\xi[X(\omega,\xi,t)X(\omega,\xi,t)^t]] - \mathbb{E}_\omega[(\mathbb{E}_\xi[X(\omega,\xi,t)])(\mathbb{E}_\xi[X(\omega,\xi,t)])^t].$$

On va s'intéresser à l'estimation de quantités plus générales, de la forme :

$$\mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))])],$$

où φ et ψ sont des fonctions régulières.

Comme présenté ci-dessus, on considère N réalisations indépendantes $(a^i)_{1\leq i\leq N}$ du champ de perméabilité a, et on note $\tilde{v}^i$ les approximations de la vitesse qui en résultent via une approximation des u^i dans l'espace d'éléments finis V_h. On définit les processus stochastiques $\tilde{X}_n^{i,j}$, pour $1 \leq i \leq N$, $1 \leq j \leq M$:

$$\left\{ \begin{array}{rcl} \tilde{X}_n^{i,j}(\omega,\xi,t_{k+1}) & = & \tilde{X}_n^{i,j}(\omega,\xi,t_k) + \tilde{v}^i(\omega,\tilde{X}_n^{i,j}(\omega,\xi,t_k))\Delta t + \sqrt{2D}(W^{i,j}(\xi,t_{k+1}) - W^{i,j}(\xi,t_k)), \\ \tilde{X}_n^{i,j}(\omega,\xi,0) & = & X_0^{i,j}(\xi), \end{array} \right. \tag{1.33}$$

où les $X_0^{i,j}$ sont des variables aléatoires indépendantes de densité $c_0(x) = \mathbb{1}_R(x)$, comme défini dans le paragraphe Advection-diffusion de polluants de la section 1.1.2, les $W^{i,j}$ sont des mouvements browniens d-dimensionnel indépendants, et où on a posé $\Delta t = T/n$ et $t_k = k\Delta t$ pour $0 \leq k \leq n$. On définit maintenant l'erreur sur l'extension moyenne généralisée :

$$E_S(\omega,\xi) = \mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))])] - \frac{1}{N}\sum_{i=1}^{N} \psi\left(\frac{1}{M}\sum_{j=1}^{M} \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))\right),$$

Théorème 1.3.5. *Sous les Hypothèses 1.3.1 et 1.3.3, pour $\varphi \in \mathcal{C}_b^3(\mathbb{R}^d,\mathbb{R}^p)$ et $\psi \in \mathcal{C}_b^1(\mathbb{R}^p,\mathbb{R}^q)$ on a l'estimation d'erreur suivante : il existe une constante C telle que*

$$\|E_S(\omega,\xi)\|_{L^2_{\omega,\xi}} \leq C\left((\Delta t)^{\frac{1+\alpha}{2}} + h|\log(h)| + \frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right).$$

On voit naturellement apparaître quatre termes : un terme de discrétisation temporelle, un terme de discrétisation spatiale et deux termes de Monte-Carlo. Nous allons donner quelques éléments de preuve pour les deux premiers, la majoration des deux derniers étant relativement immédiate. Tout d'abord notons que les vitesses approchées $\tilde{v}^i$ ne sont en général pas continues, par conséquent il faut être vigilant dans le découpage de l'erreur. On doit estimer une erreur faible de discrétisation temporelle sur l'EDS avec vitesse exacte v, c'est-à-dire avec une dérive peu régulière (plus précisément $\mathcal{C}_b^{1,\alpha}$) et avec bruit additif. Un résultat général a été obtenu dans [55], néanmoins on a ici un meilleur ordre de convergence ($\frac{1+\alpha}{2}$) au lieu de ($\frac{1}{2-\alpha}$) dans le cas d'un bruit additif. On doit ensuite estimer l'erreur sur les discrétisations temporelles provenant de l'erreur de discrétisation spatiale, on s'intéresse donc à la continuité de l'application qui à la dérive associe le processus défini par le schéma d'Euler. Cette continuité est établie avec des estimations uniformes sur les dérives, d'où la nécessité d'avoir une estimation de l'erreur éléments finis en norme $W^{1,\infty}$. Les estimations des trois premiers termes sont obtenues de manière uniforme par rapport à ω, grâce à l'Hypothèse 1.3.1. Même si nous n'avons pas été en mesure de le prouver, on s'attend, dans le cadre des hypothèses du paragraphe 1.1.2, à un résultat similaire, avec une majoration par

$$C\left((\Delta t)^{\frac{1}{4}} + h^{\frac{1}{2}-\varepsilon} + \frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right),$$

pour $\varepsilon > 0$.

On s'intéresse maintenant à l'erreur commise sur la dispersion $\mathcal{D}$ définie comme la dérivée de l'extension S. On s'intéresse donc, de manière analogue à ce qui précéde à l'estimation de quantités plus générales, de la forme

$$\frac{d}{dt}\mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))])],$$

où φ et ψ sont des fonctions régulières. On définit maintenant l'erreur sur la dispersion moyenne généralisée, par

$$\begin{aligned} E_{\mathcal{D}}(\omega,\xi) &= \frac{d}{dt}\mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))])] \\ &- \frac{1}{N}\sum_{i=1}^{N} \frac{\psi\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s)\right) - \psi\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T)\right)}{\Delta s}, \end{aligned}$$

où on a naturellement défini la valeur du processus $\tilde{X}_n^{i,j}$ en $T+\Delta s$ par

$$\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s) = \tilde{X}_n^{i,j}(\omega,\xi,T) + \tilde{v}(\tilde{X}_n^{i,j}(\omega,\xi,T))\Delta s + \sqrt{2D\Delta s}(W^{i,j}(\xi,T+\Delta s) - W^{i,j}(\xi,T)).$$

Pour estimer cette erreur qui porte sur la dérivée des quantités étudiées précédemment, on va maintenant faire une hypothèse de régularité supplémentaire.

Hypothèse 1.3.6. *On suppose que $a \in L^\infty(\Omega, \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d))$ pour un certain $\alpha > 0$, que pour presque tout ω, $a(\omega,\cdot)$ est O-périodique et vérifie pour tout $x \in \mathbb{R}^d$,*

$$0 < a^{min} \leq a(\omega,x) \leq a^{max} < +\infty.$$

On a alors l'estimation d'erreur suivante.

Théorème 1.3.7. *Sous les Hypothèses 1.3.6 et 1.3.3, pour $\varphi \in \mathcal{C}_b^5(\mathbb{R}^d,\mathbb{R}^p)$ et $\psi \in \mathcal{C}_b^2(\mathbb{R}^p,\mathbb{R}^q)$ on a l'estimation d'erreur suivante : il existe une constante C telle que*

$$\|E_{\mathcal{D}}(\omega,\xi)\|_{L^2_{\omega,\xi}} \leq C\left(\Delta t + \Delta s + h|\log(h)| + \frac{1}{\sqrt{N}} + \frac{1}{\sqrt{M\Delta s}}\right).$$

Ici, on a un découpage naturel en cinq termes : un terme de discrétisation temporelle dans le schéma d'Euler, un terme de discrétisation temporelle pour le calcul de la dérivée, un terme de discrétisation spatiale et deux termes de Monte-Carlo. On notera que pour le premier terme, on doit généraliser la notion d'erreur faible du schéma d'Euler pour une EDS au calcul de la dérivée d'une espérance. On notera que le terme en $\frac{1}{\sqrt{M\Delta s}}$ provient de l'approximation d'une dérivée par une méthode de Monte-Carlo et est optimal, comme on peut le voir dans le cas d'une EDS avec dérive constante.

Première partie

Estimations d'erreurs fortes et faibles pour des EDP elliptiques à coefficients aléatoires

Chapitre 2

Strong and weak error estimates for the solutions of elliptic partial differential equations with random coefficients

Abstract : We consider the problem of numerically approximating the solution of an elliptic partial differential equation with random coefficients and homogeneous Dirichlet boundary conditions. We focus on the case of a lognormal coefficient, we have then to deal with the lack of uniform coercivity and uniform boundedness with respect to the randomness. This model is frequently used in hydrogeology. We approximate this coefficient by a finite dimensional noise using a truncated Karhunen-Loève expansion. We give then estimates of the corresponding error on the solution, both a strong error estimate and a weak error estimate, that is to say an estimate of the error commited on the law of the solution. We obtain a weak rate of convergence wich is twice the strong one. Besides this, we give a complete error estimate for the stochastic collocation method in this case, where neither coercivity nor boundedness are stochastically uniform. To conclude, we apply these results of strong and weak convergence to two classical cases of covariance functions choices : the case of an exponential covariance kernel on a box and the case of an analytic covariance function, yielding explicit weak and strong convergence rates.

Keywords : uncertainty quantification, elliptic PDE with random coefficients, Karhunen-Loève expansion, strong error estimate, weak error estimate, lognormal distribution.

Résumé : On s'intéresse à l'approximation numérique de la solution d'une équation aux dérivées partielles elliptique à coefficients aléatoires, avec des conditions de Dirichlet homogènes au bord. On se concentre sur le cas d'un coefficient lognormal, on est ainsi confronté au fait que ce coefficient n'est ni uniformément borné, ni uniformément coercif par rapport à l'aléatoire. Ce modèle est fréquemment utilisé en hydrogéologie. On approche ce coefficient dans un espace aléatoire de dimension finie, en utilisant un développement de Karhunen-Loève. On donne alors des estimations pour l'erreur qui en découle sur la solution, une estimation d'erreur forte mais également une estimation d'erreur faible, c'est à dire une estimation de l'erreur commise sur la loi de la solution. On obtient alors un taux de convergence faible double du taux de convergence forte. De plus, on donne une estimation d'erreur complète pour la méthode de collocation ap pliquée dans ce cas où le coefficient n'est ni uniformément borné, ni uniformément coercif par rapport à l'aléatoire. Pour conclure, on applique ces résultats à deux choix particuliers de noyau de covariance : le cas d'une covariance exponentielle sur un pavé et le cas d'une covariance analytique, donnant des taux de convergence forte et faible explicites.

Mots clés : quantification des incertitudes, EDP elliptique à coefficients aléatoires, développement de Karhunen-Loève, estimation d'erreur forte, estimation d'erreur faible, distribution lognormale.

2.1 Introduction

Many engineering applications involve uncertainty on the input data, such as material properties. This uncertainty results from heterogeneity of the medium and incomplete knowledge on the medium properties and can be modelled by partial differential equations with random coefficients. This work addresses elliptic partial differential equations with random coefficients and focuses on application to hydrogeology, namely the prediction of flow in porous media, but there are many other applications, e.g., random vibrations, composite materials, seismic activity and deformations of inhomogeneous materials such as wood or biomaterials. The aim is to compute the law of the solution, but in practice we are usually interested only in some moments.
Several methods have been developed : Monte-Carlo and Monte-Carlo based methods, moment equations, perturbation methods which are adapted to the case of small uncertainty, homogenization, multiscale analysis and stochastic spectral methods [1, 2, 6, 7, 24, 27, 30, 31, 32, 46, 47, 53, 54, 58, 59, 69, 70, 71, 75], which regroup stochastic galerkin methods and stochastic collocation methods.
As in many previous works, we consider the model equation

$$-\mathrm{div}(a(\omega,x)\nabla_x u(\omega,x)) = f(\omega,x).$$

Both stochastic Galerkin methods and stochastic collocation methods are based on the approximation of a in a finite dimensional probability space, i.e. using a finite number of random variables. These methods are therefore adapted to the case where the probability space has a low dimensionality, i.e. in the case where we have a good approximation a_N of a such that a_N is a function of N random variables with N small. Such approximations a_N of a can be obtained by using either a Karhunen-Loève or a polynomial chaos expansion. We compute then the solution u_N of the approximated equation resulting from replacing a by a_N. In this paper we focus on the convergence of u_N to u.
More precisely, we work here with homogeneous Dirichlet boundary conditions and a homogeneous lognormal random field a. This is a frequently used model for flow equation in porous media. The truncated Karhunen-Loève expansion of $\log(a)$ at order N provides then an approximation a_N of a. It is important to notice that in such a case, unlike what is frequently assumed, neither the random filed a nor its approximations a_N are uniformly coercive with respect to ω. However, such elliptic PDEs with infinite dimensionnal stochastic coefficients, which are not supposed to be uniformly bounded from above and below, (typically lognormal coeeficient) have been considered in [?], [26] and in the recent work [66]. The well-posedness of the equation and Galerkin approximation have been studied in [36]. A white noise approach has been proposed in [26]. In [66], the inverse problem is adressed, using similar techniques as the ones used in our paper.
However, up to our knowledge, the strong convergence of u_N to u has never been studied under this kind of assumptions on a, and the weak convergence has never been studied. This is the main goal of this article. We first give a strong convergence result of u_N to u, i.e. a bound for the error in $L^p(H_0^1)$-norm. Then a weak convergence result is obtained, i.e. we bound the error commited on the law of u. We find a bound for the weak error whose order is twice the strong order, which presents a significant interest since the number N of random variables has to be small in order to be able to compute u_N and since the law of u is what we are interested in. For simplicity we assume that f is deterministic, but all the results can be easily extended to the case when f is random, under adequate assumptions.
To begin with, we prove the existence and uniqueness of the solution u, remarking once again that in the considered case, the random field a is not uniformly coercive with respect to ω. Then we make assumptions on the eigenvalues and on the eigenfunctions of the Karhunen-Loève expansion, which enables us to prove two preliminary results : the strong convergence of a_N to a and the existence of a uniform bound in L^p-norm for a_N^{max} and $\frac{1}{a_N^{min}}$. We can then give a strong convergence result of u_N to u, with almost sure convergence and L^p convergence. The strong error is basically bounded by the squared root of the remainder of the series of the eigenvalues. We prove next a weak convergence result, showing that the error commited on the law of u is bounded by the remainder of the series of the eigenvalues. Besides this work gives a complete convergence analysis of the collocation method, adapting the results of I. Babuska, F. Nobile and R. Tempone in [1], in which uniform coercivity of a with respect to ω is assumed. Finally we give examples of covariance kernels for which these results apply. In particular the exponential case on a box and the gaussian case are studied, these covariance functions being frequently used to model the hydraulic conductivity for the flow equation in porous media.

2.2 Equation, existence and uniqueness of the solution

In this section, we first define the homogeneous lognormal random field a and make some regularity assumptions on its covariance kernel, then we define a linear elliptic partial differential equation with random coefficients, namely a, and finally show the existence and uniqueness of the solution of this equation. Let D be an open bounded domain in $\mathbb{R}^d$ with $\mathcal{C}^2$ boundary and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We consider a function $k \in \mathcal{C}^{0,1}(\mathbb{R}, \mathbb{R})$, and $g : \Omega \times \bar{D} \to \mathbb{R}$ a mean-free gaussian field with covariance function $cov[g](x, y) = k(\|x - y\|)$.

Proposition 2.2.1. *Under these assumptions, g admits a version whose trajectories belong to $\mathcal{C}^{0,\alpha}(\bar{D})$ a.s. for $\alpha < 1/2$.*

Proof. Let us denote by L the Lipschitz constant of k.

$$\begin{aligned}\mathbb{E}[|g(x) - g(y)|^2] &= \mathbb{E}[g(x)^2] - 2\mathbb{E}[g(x)g(y)] + \mathbb{E}[g(y)^2] \\ &= 2(k(0) - k(\|x - y\|)) \\ &\leq 2L\|x - y\|.\end{aligned}$$

We recall the existence, for any positive integer p, of a constant $c_p = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^{2p} e^{-\frac{x^2}{2}} dx$ such that for all mean-free gaussian random variable X,

$$E[|X|^{2p}] \leq c_p E[|X|^2]^p.$$

Therefore, since $g(x) - g(y)$ is a mean-free gaussian random variable, we have, for any positive integer p,

$$E[|g(x) - g(y)|^{2p}] \leq c_p (2L)^p \|x - y\|^p.$$

According to the Kolmogorov continuity theorem [13], there exists a version of g which is a.s. Hölder-continuous with any exponent $\beta < \frac{p-d}{2p}$. Since this holds for any positive integer p, letting $p \to +\infty$, it follows that there exists a version of g which is a.s. Hölder continuous with any exponent $\beta < 1/2$. □

If we denote by $\tilde{g}$ this version, we clearly have that for almost all ω, $\tilde{g}(\omega, \cdot)$ and $g(\omega, \cdot)$ are equal for almost all x. If what follows we will identifie g with this version $\tilde{g}$.

We define the lognormal homogeneous random field $a : \Omega \times \bar{D} \to \mathbb{R}$ as $a(\omega, x) = e^{g(\omega,x)}$. By Proposition 2.2.1, the trajectories of a are a.s. continuous on the compact set $\bar{D}$, we can then define a.s, $a^{min}(\omega) = \min_{x \in \bar{D}} a(\omega, x)$ and $a^{max}(\omega) = \max_{x \in \bar{D}} a(\omega, x)$. We now recall Fernique theorem (see [13] and [23]), which will be next used of obtain integrability properties for a^{min} and a^{max}.

Theorem 2.2.2. *Let E be a separable Banach space, and μ a symmetric gaussian measure on $(E, \mathcal{B}(E))$. Let $\lambda > 0$ and $r > 0$ be such that*

$$\log\left(\frac{1 - \mu(\bar{B}(0, r))}{\mu(\bar{B}(0, r))}\right) + 32\lambda r^2 \leq -1.$$

Then

$$\int_E e^{\lambda\|x\|^2} \mu(dx) \leq e^{16\lambda r^2} + \frac{e^2}{e^2 - 1}.$$

From this Fernique theorem, we deduce then the following proposition.

Proposition 2.2.3. *$\frac{1}{a^{min}(\omega)} \in L^p(\Omega)$ and $a^{max}(\omega) \in L^p(\Omega)$ for any $p > 0$.*

Proof. We have $a^{min}(\omega) \geq e^{-||g(\omega)||_{\mathcal{C}^0(\bar{D})}}$ and $a^{max}(\omega) \leq e^{||g(\omega)||_{\mathcal{C}^0(\bar{D})}}$. By Fernique theorem, since g defines a mean-free gaussian measure on the Banach space $\mathcal{C}^0(\bar{D})$, there exists $\lambda > 0$ such that $\mathbb{E}[e^{\lambda||g(\omega)||^2_{\mathcal{C}^0(\bar{D})}}] < +\infty$. Thus, for any $p > 0$,

$$\mathbb{E}[e^{p||g(\omega)||_{\mathcal{C}^0(\bar{D})}}] \leq \mathbb{E}[e^{\lambda||g(\omega)||^2_{\mathcal{C}^0(\bar{D})} + \frac{p^2}{4\lambda}}] < +\infty.$$

Therefore $e^{||g(\omega)||_{\mathcal{C}^0(\bar{D})}} \in L^p(\Omega)$, and finally $\frac{1}{a^{min}(\omega)} \in L^p(\Omega)$ and $a^{max}(\omega) \in L^p(\Omega)$, for any $p > 0$. □

Proposition 2.2.4. *Let f in $L^2(D)$, then the equation :*

$$\begin{cases} -\mathrm{div}(a(\omega,x)\nabla u(\omega,x)) &= f(x) \quad \text{on } D, \\ u &= 0 \quad \text{on } \partial D, \end{cases}$$

admits a unique solution u, which belongs to $L^p(\Omega, H_0^1(D))$, for any $p > 0$.

Remark 2.2.5. *All the following results hold in the case where the forcing term f is stochastic, under adequate assumptions.*

Proof. For almost all $\omega \in \Omega$, the equation admits a unique solution $u(\omega) \in H_0^1(D)$, the mapping $\omega \mapsto u(\omega)$ is measurable and we have, a.s. :

$$\|u(\omega,x)\|_{H_0^1(D)} \leq C_D \frac{\|f\|_{L^2(D)}}{a^{min}(\omega)} \tag{2.1}$$

where C_D is the constant given by Poincaré inequality. For every $p > 0$, by Proposition 2.2.3, $\frac{1}{a^{min}} \in L^p(\Omega)$, so for any $p > 0$,

$$\mathbb{E}\left[\|u(\omega,x)\|^p_{H_0^1(D)}\right] \leq C_D^p \mathbb{E}\left[\left(\frac{1}{a^{min}(\omega)}\right)^p\right] \|f\|^p_{L^2(D)} < +\infty.$$

□

2.3 Strong convergence of a_N to a

In this section, we define the approximated random field a_N and study the strong convergence of a_N to a, i.e. the $L^p(\Omega, \mathcal{C}^0(\bar{D}))$-convergence, and the almost-sure convergence. Let $\{(\lambda_n, b_n)\}$ denote the sequence of eigenpairs associated with the compact self-adjoint operator that maps

$$f \in L^2(D) \mapsto \int_D \mathrm{cov}[g](x,.)f(x)dx \ \in L^2(D),$$

where $\lambda_1 \geq \lambda_2 \geq ... \geq 0$, and the eigenfunctions are orthonormal. We recall that $\mathbb{E}[g(\omega,x)] = 0$ and that $\mathrm{cov}[g](x,y) = k(\|x-y\|)$. Therefore $\sum_{n\geq 1} \lambda_n = |D|k(0)$. Then the truncated Karhunen-Loève expansion [49, 50] g_N of the stochastic gaussian process g and its exponential a_N are defined by

$$g_N(\omega,x) = \sum_{n=1}^{N} \sqrt{\lambda_n} b_n(x) Y_n(\omega), \quad a_N(\omega,x) = e^{\sum_{n=1}^N \sqrt{\lambda_n} b_n(x) Y_n(\omega)}.$$

where the real random variables $(Y_n)_{n\geq 1}$ are uniquely determined by

$$Y_n(\omega) = \frac{1}{\sqrt{\lambda_n}} \int_D g(\omega,x) b_n(x) dx.$$

They are independent gaussian random variables with mean zero and unit variance. Mercer theorem [60] gives the following convergence result :

$$\sup_{x\in D} \mathbb{E}[(g-g_N)^2](x) \to 0 \text{ as } N \to +\infty.$$

This convergence result does not actually enable us to conclude on the convergence of the solution.
From now on, we make the following assumption :

Assumption 2.3.1. *i) The eigenfunctions b_n are continuously differentiable.*
ii) The series $\sum_{n\geq 1} \lambda_n \|b_n\|^2_\infty$ is convergent.

iii) There exists $0 < \alpha < 1$ *such that* $\sum_{n\geq 1} \lambda_n \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha}$ *is convergent.*

Remark 2.3.2. *Such assumptions are fulfilled in the case of an exponential covariance kernel on a rectangular domain (with any* $\alpha < 1/2$*), and in the case of a gaussian covariance kernel, and more generally in the case of an analytic covariance function, for more details see section .*

Definition 2.3.3. *In the case where this assumption is fulfilled, we define*

$$R_N^\alpha = \max\left(\sum_{n>N} \lambda_n \|b_n\|_\infty^2, \sum_{n>N} \lambda_n \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha}\right),$$

for $0 \leq \alpha < 1$ *such that the second series is convergent.*

Proposition 2.3.4. *For any* α *as in Assumption 2.3.1,*$\beta < \alpha$ *and* $p > 0$*, there exists a constant* $A_{\alpha,\beta,p}$ *such that for all* N *in* $\mathbb{N}$ *:*

$$\|g_N - g\|_{L^p(\Omega, \mathcal{C}^{0,\beta}(\bar{D}))} \leq A_{\alpha,\beta,p} (R_N^\alpha)^{\frac{1}{2}}.$$

In particular, for any α *as in Assumption 2.3.1 and* $p > 0$*, there exists a constant* $A_{\alpha,p}$ *such that for all* N *in* $\mathbb{N}$*,*

$$\|g_N - g\|_{L^p(\Omega, \mathcal{C}^{0}(\bar{D}))} \leq A_{\alpha,p} (R_N^\alpha)^{\frac{1}{2}}.$$

Proof. For all $x, y \in D$, and α as in Assumption 2.3.1,

$$\begin{aligned} (b_n(x) - b_n(y))^2 &\leq (2\|b_n\|_\infty)^{2(1-\alpha)} (\|\nabla b_n\|_\infty \|x-y\|)^{2\alpha} \\ &\leq 4\|x-y\|^{2\alpha} \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha}. \end{aligned} \tag{2.2}$$

From this we deduce that

$$\begin{aligned} \mathbb{E}\left[((g_N - g)(x) - (g_N - g)(y))^2\right] &= \sum_{n>N} \lambda_n (b_n(x) - b_n(y))^2 \\ &\leq 4\left(\sum_{n>N} \lambda_n \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha}\right) \|x-y\|^{2\alpha} \\ &\leq 4R_N^\alpha \|x-y\|^{2\alpha}. \end{aligned}$$

For all $x, y \in D$ and N in $\mathbb{N}$, $(g_N - g)(x) - (g_N - g)(y)$ is a mean-free gaussian random variable, as a limit in L^2 of a linear combination of independent gaussian random variables,

$$(g_N - g)(x) - (g_N - g)(y) = \lim_{p\to+\infty} \sum_{n=N+1}^{p} \sqrt{\lambda_n} Y_n(\omega)(b_n(x) - b_n(y)).$$

Arguing as in the proof of Proposition 2.2.1 yields

$$\begin{aligned} \mathbb{E}\left[((g_N - g)(x) - (g_N - g)(y))^{2p}\right] &\leq c_p \mathbb{E}\left[((g_N - g)(x) - (g_N - g)(y))^2\right]^p \\ &\leq 4^p c_p (R_N^\alpha)^p \|x-y\|^{2\alpha p}. \end{aligned}$$

Thus, for any α as in Assumption 2.3.1, for any positive integer p, there exists a constant C_p such that :

$$\mathbb{E}\left[((g_N - g)(x) - (g_N - g)(y))^{2p}\right] \leq C_p (R_N^\alpha)^p \|x-y\|^{2\alpha p}.$$

We now are going to use the proof of the Kolmogorov continuity theorem based on Sobolev embeddings, which can be found in [13]. For any α as in Assumption 2.3.1, let ν be such that $2\alpha p - d - 2p\nu > -d$ i.e.

$\nu < \alpha$, then using Fubini theorem and the prevous bound, we get :

$$\begin{aligned}
\mathbb{E}[\|g_N - g\|^{2p}_{W^{\nu,2p}}] &= \int_D \mathbb{E}[(g_N - g)^{2p}(x)]dx + \int_\Omega \int_D \int_D \frac{|(g_N - g)(x) - (g_N - g)(y)|^{2p}}{\|x - y\|^{d+2p\nu}} dxdydP(\omega) \\
&\leq c_p \int_D \mathbb{E}[(g - g_N)^2(x)]^p dx + C_p (R_N^\alpha)^p \int_D \int_D \|x - y\|^{2\alpha p - d - 2p\nu} dxdy \\
&\leq c_p \int_D (\sum_{n>N} \lambda_n b_n(x)^2)^p dx + C_p (R_N^\alpha)^p \int_D \int_D \|x - y\|^{2\alpha p - d - 2p\nu} dxdy \\
&\leq c_p |D| (R_N^\alpha)^p + C_p (R_N^\alpha)^p \int_D \int_D \|x - y\|^{2\alpha p - d - 2p\nu} dxdy \\
&\leq C_{\alpha,p,\nu} (R_N^\alpha)^p.
\end{aligned}$$

with $C_{\alpha,p,\nu} = c_p|D| + C_p \int_D \int_D \|x - y\|^{2\alpha p - d - 2p\nu} dxdy < +\infty$.
Next we use that $W^{\nu,2p}$ is continuously embedded in $\mathcal{C}^{0,\beta}(\bar{D})$ for $\beta < \min\{1, \nu - \frac{d}{2p}\}$, let $k_{\beta,\nu,p}$ denotes the norm of this continuous embedding. For any α as in Assumption 2.3.1, for any β such that $\beta < \alpha$, there exists $p_0 \geq 1$ such that for all $p \geq p_0$, $\beta + \frac{d}{2p} < \alpha$, then we choose ν such that $\beta + \frac{d}{2p_0} < \nu < \alpha$, finally for $p \geq p_0$, $g_N - g \in \mathcal{C}^{0,\beta}(\bar{D})$ for almost all ω, with :

$$\mathbb{E}[\|g_N - g\|^{2p}_{\mathcal{C}^{0,\beta}(\bar{D})}] \leq k^{2p}_{\beta,\nu,p} C_{\alpha,p,\nu} (R_N^\alpha)^p.$$

Therefore, for any α as in Assumption 2.3.1, for any β such that $\beta < \alpha$, there exists $p_0 > 0$ such that for any $p \geq p_0$, there exists a constant $\tilde{C}_{\alpha,\beta,p}$ such that for any N in $\mathbb{N}$:

$$\|g_N - g\|_{L^{2p}(\Omega, \mathcal{C}^{0,\beta}(\bar{D}))} \leq \tilde{C}_{\alpha,\beta,p} (R_N^\alpha)^{\frac{1}{2}}.$$

Since in a probability space, if $p \leq q$, $f \in L^q$ implies $f \in L^p$ with $\|f\|_{L^p} \leq \|f\|_{L^q}$, we can conclude that for any $p > 0$ and α as in Assumption 2.3.1, for any β such that $\beta < \alpha$, there exists a constant $A_{\alpha,\beta,p}$ such that for all N in $\mathbb{N}$:

$$\|g_N - g\|_{L^p(\Omega, \mathcal{C}^{0,\beta}(\bar{D}))} \leq A_{\alpha,\beta,p} (R_N^\alpha)^{\frac{1}{2}}.$$

□

To get almost sure convergence, we need an additionnal assumption.

Assumption 2.3.5. *We still suppose that Assumption 2.3.1 is fulfilled and consider a corresponding α, we then suppose that there exists $p_0 > 0$ such that the series*

$$\sum_{N>0} (R_N^\alpha)^{p_0}$$

is convergent.

Proposition 2.3.6. *We suppose that Assumption 2.3.5 is fulfilled and consider corresponding α and p_0. Then, for any $\beta < \alpha$, for almost all ω, $g_N \to g$ in $\mathcal{C}^{0,\beta}(\bar{D})$ as $N \to +\infty$ and so $a_N \to a$ in $\mathcal{C}^0(\bar{D})$ as $N \to +\infty$.*
It follows that a_N^{max} converges a.s. to a^{max} and a_N^{min} converges a.s. to a^{min} as $N \to +\infty$.

Proof. We use the Borel-Cantelli lemma : by the previous proposition, there exists a constant $A_{\alpha,\beta,2p_0}$ such that for all N in $\mathbb{N}$:

$$\|g_N - g\|_{L^{2p_0}(\Omega, \mathcal{C}^{0,\beta}(\bar{D}))} \leq A_{\alpha,\beta,2p_0} (R_N^\alpha)^{\frac{1}{2}},$$

which implies by Markov inequality that for any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}(\|g_N - g\|_{C^{0,\beta}(\bar{D})} \geq \varepsilon) &\leq \frac{\|g_N - g\|^{2p_0}_{L^{2p_0}(\Omega, C^{0,\beta}(\bar{D}))}}{\varepsilon^{2p_0}} \\ &\leq \frac{A^{2p_0}_{\alpha,\beta,2p_0}(R_N^\alpha)^{p_0}}{\varepsilon^{2p_0}}.\end{aligned}$$

Therefore we have $\sum_{N\geq 1} \mathbb{P}(\|g_N - g\|_{C^{0,\beta}} \geq \varepsilon) < +\infty$. The Borel-Cantelli lemma yields $\mathbb{P}(\limsup(\|g_N - g\|_{C^{0,\beta}} \geq \varepsilon)) = 0$ for all $\varepsilon > 0$, and so $g_N \to g$ a.s. in $\mathcal{C}^{0,\beta}(\bar{D})$ as $N \to +\infty$. Finally, thanks to the continuity of the exponential function, we conclude that $a_N = e^{g_N} \to a = e^g$ a.s. in $\mathcal{C}^0(\bar{D})$ as $N \to +\infty$. □

The following results, which give uniform bounds for the random variables a_N^{max} and $\frac{1}{a_N^{min}}$ (and more general quantities, which will be used in section) in the L^p-norm will be used to conclude this section and in the next two sections.

Definition 2.3.7. *For $N \in \mathbb{N}$, $\gamma \in [0,1]^N$, we define $g_{\gamma,N} : \Omega \times D \to \mathbb{R}$ and $a_{\gamma,N} : \Omega \times D \to \mathbb{R}$ by : $g_{\gamma,N}(\omega, x) = \sum_{n=1}^N \sqrt{\lambda_n} Y_n(\omega)\gamma_n b_n(x)$, and $a_{\gamma,N}(\omega, x) = e^{g_{\gamma,N}(\omega,x)}$.*

Proposition 2.3.8. *For any α as in Assumption 2.3.1, $\beta < \alpha$ and $p > 0$, there exists a constant $B_{\beta,p}$ such that for any N in $\mathbb{N}$, and γ in $[0,1]^N$.*

$$||g_{\gamma,N}||_{L^p(\Omega, C^{0,\beta}(\bar{D}))} \leq B_{\beta,p}$$

Proof. Using similar techniques as in the proof of Proposition 2.3.4, we have by inequality (2.2) that, for any $N \in \mathbb{N}$, $\gamma \in [0,1]^N$

$$\begin{aligned}\mathbb{E}[(g_{\gamma,N}(\omega, x) - g_{\gamma,N}(\omega, y))^2] &= \sum_{n=1}^N \lambda_n \gamma_n^2 (b_n(x) - b_n(y))^2 \\ &\leq 4\left(\sum_{n=1}^N \lambda_n \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha}\right) \|x - y\|^{2\alpha} \\ &\leq 4R_0^\alpha \|x - y\|^{2\alpha}.\end{aligned}$$

Since, for any x, y in D, N in $\mathbb{N}$, $(g^{\gamma,N}(x) - g^{\gamma,N}(y))$ is a mean-free gaussian random variable, we have :

$$\mathbb{E}[(g_{\gamma,N}(\omega, x) - g_{\gamma,N}(\omega, y))^{2p}] \leq 4^p c_p (R_0^\alpha)^p \|x - y\|^{2\alpha p}.$$

Therefore, for any α as in Assumption 2.3.1, for any $p \geq 1$, there exists a constant $M_{\alpha,p}$ such that, for any $N \in \mathbb{N}$ and $\gamma \in [0,1]^N$, we have :

$$\mathbb{E}[(g_{\gamma,N}(\omega, x) - g_{\gamma,N}(\omega, y))^{2p}] \leq M_{\alpha,p} \|x - y\|^{2\alpha p},$$

where $M_{\alpha,p} = c_p (R_0^\alpha)^p$. Then, for $\nu < \alpha$ and p such that we have

$$\mathbb{E}[\|g_{\gamma,N}\|^{2p}_{W^{\nu,2p}}] \leq c_p |D| (R_0^\alpha)^p + M_{\alpha,p} \int_D \int_D \|x - y\|^{2\alpha p - d - 2p\nu} dx dy.$$

We finally conclude as in the proof of Proposition 2.3.4. □

Definition 2.3.9. *By Assumption 2.3.1, for any $N \in \mathbb{N}$ and $\gamma \in [0,1]^N$, the trajectories of $a_{\gamma,N}$ are continuous on the compact set $\bar{D}$ a.s, so we can define, for almost all $\omega \in \Omega$, $a^{max}_{\gamma,N}(\omega) = \max_{x\in\bar{D}} a_{\gamma,N}(\omega, x)$ and $a^{min}_{\gamma,N}(\omega) = \min_{x\in\bar{D}} a_{\gamma,N}(\omega, x)$.*

We can finally bound $a^{max}_{\gamma,N}$ and $\frac{1}{a^{min}_{\gamma,N}}$ in L^p-norm, independently from N and γ.

Proposition 2.3.10. *For any $p>0$, $a^{max}_{\gamma,N}$ and $\frac{1}{a^{min}_{\gamma,N}} \in L^p(\Omega)$, and there exists a constant D_p such that for any $N\in\mathbb{N}$, and $\gamma\in[0,1]^N$*

$$\left\| \frac{1}{a^{min}_{\gamma,N}} \right\|_{L^p(\Omega)} \leq D_p, \text{ and } \quad \|a^{max}_{\gamma,N}\|_{L^p(\Omega)} \leq D_p.$$

In particular,

$$\|a_N\|_{L^p(\Omega,\mathcal{C}^0(\bar{D}))} \leq D_p.$$

Proof. We apply Fernique Theorem 2.2.3, uniformly with respect to N and γ. There exists $x_0\in]0,1[$ such that for all $x\in[x_0,1[$, $\log\left(\frac{1-x}{x}\right)\leq -2$. Proposition 2.3.8 yields the existence of a constant B_2 such that for any $N\in\mathbb{N}$, and $\gamma\in[0,1]^N$:

$$||g_{\gamma,N}||_{L^2(\Omega,\mathcal{C}^0(\bar{D}))} \leq B_2.$$

Thus, setting $r_0=\frac{B_2}{\sqrt{1-x_0}}$, we have, for every $r\geq r_0$,

$$\begin{aligned}\mathbb{P}(\|g_{\gamma,N}\|_{\mathcal{C}^0(\bar{D})}\geq r) &\leq \frac{\|g_{\gamma,N}\|^2_{L^2(\Omega,\mathcal{C}^0(\bar{D}))}}{r^2}\\ &\leq \frac{B_2^2}{r^2}\leq 1-x_0.\end{aligned}$$

We now choose λ such that $32\lambda r_0^2\leq 1$, and we denote by $\mu_{\gamma,N}$ the law of $g_{\gamma,N}:\Omega\to\mathcal{C}^0(\bar{D})$. Since the $\mu_{\gamma,N}$ are centred gaussian measures on the Banach space $\mathcal{C}^0(\bar{D})$, we have then, for any $N\in\mathbb{N}$, and $\gamma\in[0,1]^N$,

$$\log\left(\frac{1-\mu_{\gamma,N}(\bar{B}(0,r_0))}{\mu_{\gamma,N}(\bar{B}(0,r_0))}\right)+32\lambda r_0^2\leq -1.$$

Use Fernique Theorem 2.2.3, set $k=e^{16\lambda r_0^2}+\frac{e^2}{e^2-1}$, to obtain that for all $N\in\mathbb{N}$, and $\gamma\in[0,1]^N$,

$$\mathbb{E}[e^{\lambda\|g_{\gamma,N}\|^2_{\mathcal{C}^0(\bar{D})}}]\leq k.$$

Hence, for any $p>0$, $N\in\mathbb{N}$, and $\gamma\in[0,1]^N$,

$$\begin{aligned}\mathbb{E}[e^{p\|g_{\gamma,N}\|_{\mathcal{C}^0(\bar{D})}}] &\leq e^{\frac{p^2}{4\lambda}}\mathbb{E}[e^{\lambda\|g_{\gamma,N}\|^2_{\mathcal{C}^0(\bar{D})}}]\\ &\leq ke^{\frac{p^2}{4\lambda}}.\end{aligned}$$

Denoting $D_p=(ke^{\frac{p^2}{4\lambda}})^{\frac{1}{p}}$, we conclude that :

$$\left\| \frac{1}{a^{min}_{\gamma,N}} \right\|_{L^p(\Omega)} \leq \|e^{\|g_{\gamma,N}\|_{\mathcal{C}^0(\bar{D})}}\|_{L^p(\Omega)}\leq D_p,$$

and

$$\|a^{max}_{\gamma,N}\|_{L^p(\Omega)} \leq \|e^{\|g_{\gamma,N}\|_{\mathcal{C}^0(\bar{D})}}\|_{L^p(\Omega)}\leq D_p.$$

□

Proposition 2.3.11. *For any $p>0$, and α as in Assumption 2.3.1, there exists a constant $E_{\alpha,p}$ such that for any $N\in\mathbb{N}$,*

$$\|a_N-a\|_{L^p(\Omega,\mathcal{C}^0(\bar{D}))}\leq E_{\alpha,p}(R_N^{\alpha})^{\frac{1}{2}}$$

Proof. Take $p > 0$, choose $q, r > 0$ such that $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, then the following inequality

$$\forall x, y \in \mathbb{R} \quad |e^x - e^y| \leq |x - y|(e^x + e^y),$$

together with Hölder's inequality leads to :

$$\|e^{g_N} - e^g\|_{L^r(\Omega, \mathcal{C}^0(\bar{D}))} \leq \|g_N - g\|_{L^p(\Omega, \mathcal{C}^0(\bar{D}))} \|e^{g_N} + e^g\|_{L^q(\Omega, \mathcal{C}^0(\bar{D}))},$$

which we rewrite as

$$\begin{aligned} \|a_N - a\|_{L^r(\Omega, \mathcal{C}^0(\bar{D}))} &\leq \|g_N - g\|_{L^p(\Omega, \mathcal{C}^0(\bar{D}))} \|a_N + a\|_{L^q(\Omega, \mathcal{C}^0(\bar{D}))} \\ &\leq A_{\alpha,p} (R_N^\alpha)^{\frac{1}{2}} (D_q + \|a\|_{L^q(\Omega, \mathcal{C}^0(\bar{D}))}). \end{aligned}$$

We conclude by setting $E_{\alpha,r} = A_{\alpha,p}(D_q + \|a\|_{L^q(\Omega, \mathcal{C}^0(\bar{D}))})$, with $p = q = 2r$ for instance. □

2.4 Strong convergence of u_N to u

Thanks to the results of the previous section, we can now estimate the strong error commited on the solution u, resulting from the approximation of a by a_N.

Since for all $N \in \mathbb{N}$ and $\gamma \in [0,1]^N$, the random variables $a_{\gamma,N}^{max}$ and $\frac{1}{a_{\gamma,N}^{min}}$ belong to $L^p(\Omega)$ for all $p > 0$, the equation

$$\left\{ \begin{array}{rll} -\mathrm{div}(a_{\gamma,N}(\omega, x) \nabla u_N^\gamma(\omega, x)) &= f(x) & \text{on } D, \\ u_N^\gamma(\omega, x) &= 0 & \text{on } \partial D, \end{array} \right. \tag{2.3}$$

admits a unique solution $u_N^\gamma \in L^p(\Omega, H_0^1(D))$ for all $p > 0$. In particular, for $\gamma = 1^N$, $a_{\gamma,N} = a_N$ and we denote by u_N the solution. Let us set for $(y_1, ..., y_N) \in \mathbb{R}^N$ and $x \in D$, $\tilde{a}_N(y_1, ..., y_N, x) = e^{\sum_{n=1}^N \sqrt{\lambda_i} b_i(x) y_i}$ and $\tilde{u}_N(y_1, ..., y_N, .)$ be the solution of

$$\left\{ \begin{array}{rll} -\mathrm{div}_x(\tilde{a}_N(y, x) \nabla_x \tilde{u}_N(y, x)) &= f(x) & \text{on } D, \\ \tilde{u}_N(y, x) &= 0 & \text{on } \partial D. \end{array} \right. \tag{2.4}$$

It is classical that $\tilde{u}_N$ is a $\mathcal{C}^\infty$ function of $y_1, ..., y_N$. When we need to emphasize the dependance of $\tilde{u}_N$ on $y_1, ..., y_N$, we write $\tilde{u}_N(y_1, ..., y_N)$. We have then $\tilde{u}_N \in \mathcal{C}^\infty(\mathbb{R}^N, H_0^1(D))$. We notice that a.s. $a_N(\omega, x) = \tilde{a}_N(Y_1(\omega), ..., Y_N(\omega), x)$, and $u_N(\omega, x) = \tilde{u}_N(Y_1(\omega), ..., Y_N(\omega), x)$. For convenience, $\tilde{a}_N$ will still be denoted by a_N and $\tilde{u}_N$ by u_N .

We first show the almost sure convergence of u_N to u.

Proposition 2.4.1. *Under Assumption 2.3.5, $u_N(\omega, x)$ converges to $u(\omega, x)$ in $H_0^1(D)$, for almost all ω.*

Proof. By Proposition 2.3.6, for almost all ω, a_N converges to a in $\mathcal{C}^0(\bar{D})$ i.e. uniformly. Then we use the continuity of the solution u with respect to the coefficient a of the equation, indeed we have a.s. :

$$\begin{aligned} a_N^{min} \|u - u_N\|_{H_0^1(D)}^2 &\leq \int_D a_N |\nabla(u - u_N)|^2 dx \\ &= \int_D (a_N - a) \nabla u \nabla(u - u_N) dx \\ &\leq \|a - a_N\|_{\mathcal{C}^0(\bar{D})} \|u - u_N\|_{H_0^1(D)} \|u\|_{H_0^1(D)}. \end{aligned}$$

Therefore, thanks to (2.1), we have, for almost all ω :

$$\|u - u_N\|_{H_0^1(D)} \leq \frac{1}{a_N^{min}} \|a - a_N\|_{\mathcal{C}^0(\bar{D})} \|f\|_{L^2(D)} \frac{C_D}{a^{min}}.$$

The right-hand side of this inequality converges a.s. to 0 as $N \to +\infty$ by Proposition 2.3.6. □

Next we give a convergence result and an error estimate in L^p-norm.

Theorem 2.4.2. *For all $p > 0$, u_N converges to u in $L^p(\Omega, H_0^1(D))$, and for any α as in Assumption 2.3.1, there exists a constant $F_{\alpha,p}$ such that*

$$\|u - u_N\|_{L^p(\Omega, H_0^1(D))} \leq F_{\alpha,p}(R_N^\alpha)^{\frac{1}{2}}.$$

Proof. For all $p > 0$, for almost all ω,

$$\|u - u_N\|_{H_0^1(D)} \leq \frac{1}{a_N^{min}} \|a - a_N\|_{\mathcal{C}^0(\bar{D})} \|f\|_{L^2} \frac{C_D}{a^{min}}.$$

Hence, by choosing $q, r, s > 0$ such that $\frac{1}{p} = \frac{1}{q} + \frac{1}{r} + \frac{1}{s}$ it follows from Hölder inequality and Proposition 2.3.11 that for any $0 < \alpha \leq b$

$$\begin{aligned} \|u - u_N\|_{L^p(\Omega, H_0^1(D))} &\leq \left\| \frac{1}{a_N^{min}} \right\|_{L^q(\Omega)} \|a - a_N\|_{L^r(\Omega, \mathcal{C}^0)} \|f\|_{L^2} C_D \left\| \frac{1}{a^{min}} \right\|_{L^s(\Omega)} \\ &\leq D_q \|f\|_{L^2} D_s E_{\alpha,r} (R_N^\alpha)^{\frac{1}{2}}. \end{aligned}$$

□

The following results gives a bound for u_N^γ in $L^p(H_0^1)$-norm independent of N and γ, which will be useful in the next section.

Lemma 2.4.3. *For all $p > 0$, there exists a constant G_p such that for all $N \in \mathbb{N}$, and $\gamma \in [0,1]^N$*

$$\|u_N^\gamma\|_{L^p(\Omega, H_0^1(D))} \leq G_p.$$

Proof. For almost all ω, for any $p > 0$, $N \in \mathbb{N}$, and $\gamma \in [0,1]^N$, we have by Proposition 2.3.10 :

$$\begin{aligned} \|u_N^\gamma\|_{H_0^1(D)} &\leq \frac{C_D}{a_{min}^{\gamma,N}} \|f\|_{L^2(D)}, \\ \|u_N^\gamma\|_{L^p(\Omega, H_0^1(D))} &\leq C_D \left\| \frac{1}{a_{min}^{\gamma,N}} \right\|_{L^p} \|f\|_{L^2(D)} \\ &\leq C_D D_p \|f\|_{L^2(D)}. \end{aligned}$$

□

2.5 Weak convergence of u_N to u

In this section we are interested in the error committed on the law of u, more precisely we show that the order of the bound for the weak convergence of u_N to u is twice the order of the bound for the strong convergence. In order to estimate the weak error, that is to say the expected value of $\varphi(u_N) - \varphi(u)$, for some regular function φ, we need estimates on the growth of the derivatives of $\varphi(u_N)$ with respect to the y_i, which follow from the following estimates on the derivatives of u_N with respect to the y_i. In this section, we only consider the spatial dimension $d = 1$, 2 or 3.

Proposition 2.5.1. *For any integer k, there exists a constant $C(k)$ such that for any $N \in \mathbb{N}$, for any multi-index $\alpha \in \mathbb{N}^N$ with length k, we have the following estimate on the growth of the derivatives of u_N with respect to y :*

$$\left\| \frac{\partial^\alpha u_N}{\partial y^\alpha} \right\|_{H_0^1(D)} \leq C(k) \sqrt{\frac{a_{max}^N(y)}{a_{min}^N(y)}} ||u_N||_{H_0^1} \prod_{i \in \mathbb{N}} \sqrt{\lambda_i^{\alpha_i}} \|b_n\|_\infty^{\alpha_i}.$$

Proof. We recall that for all $y \in \mathbb{R}^N$, $u(y,.)$ solves the following equation

$$\int_D a_N(y,x)\nabla_x u_N(y,x)\nabla_x v(x) = \int_D f(x)v(x), \quad \forall v \in H_0^1(D).$$

We compute the derivatives of a_N with respect to the y_i. For all $1 \leq i,j \leq N$:

$$\frac{\partial a_N}{\partial y_i}(y,x) = \sqrt{\lambda_i}b_i(x)a_N(y,x), \quad \frac{\partial a_N}{\partial y_i \partial y_j}(y,x) = \sqrt{\lambda_i}b_i(x)\sqrt{\lambda_j}b_j(x)a_N(y,x).$$

In every point $y \in \mathbb{R}^N$, the derivatives of u with respect to y_i and with respect to y_i and y_j, for $1 \leq i,j \leq N$ satisfy : $\forall v \in H_0^1(D)$

$$\int_D \frac{\partial a_N}{\partial y_i}(y,x)\nabla u_N(y,x)\nabla v(x) + \int_D a_N(y,x)\nabla\frac{\partial u_N}{\partial y_i}(y,x)\nabla v(x) = 0,$$
$$\int_D \left(\frac{\partial^2 a_N}{\partial y_i \partial y_j}\nabla u_N + \frac{\partial a_N}{\partial y_i}\nabla\frac{\partial u_N}{\partial y_j} + \frac{\partial a_N}{\partial y_j}\nabla\frac{\partial u_N}{\partial y_i} + a_N\nabla\frac{\partial^2 u_N}{\partial y_i \partial y_j}\right)(y,x)\nabla v(x) = 0.$$

Choosing $v(x) = \frac{\partial u_N}{\partial y_i}(y,x)$ in the first variational formulation, we have :

$$\begin{aligned}\left\|\sqrt{a_N}(y,x)\nabla\frac{\partial u_N}{\partial y_i}\right\|_{L^2} &\leq \sqrt{\lambda_i}\|b_i\|_\infty||\sqrt{a_N}(y,x)\nabla u_N||_{L^2} \\ &\leq \sqrt{\lambda_i}\|b_i\|_\infty\sqrt{a^N_{max}(y)}||u_N||_{H_0^1}, \\ \left\|\frac{\partial u_N}{\partial y_i}\right\|_{H_0^1} &\leq \sqrt{\lambda_i}\|b_i\|_\infty\sqrt{\frac{a^N_{max}(y)}{a^N_{min}(y)}}||u_N||_{H_0^1}.\end{aligned}$$

Choosing $v(x) = \frac{\partial^2 u_N}{\partial y_i \partial y_j}(y,x)$ in the second variational formulation, we obtain :

$$\begin{aligned}\left\|\sqrt{a_N}(y,x)\nabla\frac{\partial^2 u_N}{\partial y_i \partial y_j}\right\|_{L^2} &\leq \sqrt{\lambda_i}\sqrt{\lambda_j}\|b_i\|_\infty\|b_j\|_\infty||\sqrt{a_N}\nabla u_N||_{L^2} \\ &+ \sqrt{\lambda_i}\|b_i\|_\infty||\sqrt{a_N}\nabla\frac{\partial u_N}{\partial y_j}||_{L^2} + \sqrt{\lambda_j}\|b_j\|_\infty||\sqrt{a_N}\nabla\frac{\partial u_N}{\partial y_i}||_{L^2} \\ &\leq 3\sqrt{a^N_{max}(y)}\sqrt{\lambda_i}\sqrt{\lambda_j}\|b_i\|_\infty\|b_j\|_\infty||u_N||_{H_0^1}.\end{aligned}$$

Therefore :

$$\left\|\frac{\partial^2 u_N}{\partial y_i \partial y_j}(y)\right\|_{H_0^1} \leq 3\sqrt{\frac{a^N_{max}(y)}{a^N_{min}(y)}}\sqrt{\lambda_i}\sqrt{\lambda_j}\|b_i\|_\infty\|b_j\|_\infty||u_N||_{H_0^1}.$$

The result follows for $|\alpha| \geq 3$ by induction. □

Proposition 2.5.2. *Let $\varphi \in \mathcal{C}^4(\mathbb{R},\mathbb{R})$, whose derivatives are bounded by a constant C_φ, then for any $N \in \mathbb{N}$, for any multi-index $\alpha \in \mathbb{N}^N$ with $|\alpha| \leq 4$, there exists a constant $C'(|\alpha|,p)$ such that we have the following estimate on the growth of the derivatives of $\varphi \circ u_N$ with respect to y :*

$$\left\|\frac{\partial^\alpha \varphi \circ u_N}{\partial y^\alpha}\right\|_{L^p(D)} \leq C'(|\alpha|,p)C_\varphi\left(1 + \sqrt{\frac{a^N_{max}(y)}{a^N_{min}(y)}}||u_N||_{H_0^1}\right)^{|\alpha|}\prod_{i\in\mathbb{N}}\sqrt{\lambda_i^{\alpha_i}}\|b_i\|_\infty^{\alpha_i},$$

where $p \leq \infty$ if $d = 1$, $p < \infty$ if $d = 2$, and $p \leq \frac{3}{2}$ if $d = 3$ and $C'(|\alpha|,p)$ only depends on the length of α and p.

Proof. We first note that under our assumption, by Sobolev embedding we know that $H_0^1(D) \subset L^r(D)$ for $r \leq 4p$, we then denote by κ_r the norm of this Sobolev embedding.

For $|\alpha| = 1$, we have :

$$\begin{aligned} \frac{\partial \varphi \circ u_N}{\partial y_i}(y) &= \varphi' \circ u_N(y) \frac{\partial u_N}{\partial y_i}(y) \\ \left\| \frac{\partial \varphi \circ u_N}{\partial y_i}(y) \right\|_{L^p} &\leq C_\varphi \left\| \frac{\partial u_N}{\partial y_i}(y) \right\|_{L^p} \\ &\leq \kappa_p C_\varphi \sqrt{\lambda_i} \|b_i\|_\infty \sqrt{\frac{a_{max}^N(y)}{a_{min}^N(y)}} ||u_N||_{H_0^1}. \end{aligned}$$

Then, for $|\alpha| = 2$, using Hölder inequality :

$$\begin{aligned} \frac{\partial^2 \varphi \circ u_N}{\partial y_i \partial y_j}(y) &= \varphi' \circ u_N(y) \frac{\partial^2 u_N}{\partial y_i \partial y_j}(y) + \varphi''(u_N)(y) \frac{\partial u_N}{\partial y_i}(y) \frac{\partial u_N}{\partial y_j}(y) \\ \left\| \frac{\partial^2 \varphi \circ u_N}{\partial y_i \partial y_j}(y) \right\|_{L^p} &\leq C_\varphi \left\| \frac{\partial^2 u_N}{\partial y_i \partial y_j}(y) \right\|_{L^p} + C_\varphi \left\| \frac{\partial u_N}{\partial y_i}(y) \right\|_{L^{2p}} \left\| \frac{\partial u_N}{\partial y_j}(y) \right\|_{L^{2p}} \\ &\leq \kappa_p C_\varphi \left\| \frac{\partial^2 u_N}{\partial y_i \partial y_j}(y) \right\|_{H_0^1} + \kappa_{2p} C_\varphi \left\| \frac{\partial u_N}{\partial y_i}(y) \right\|_{H_0^1} \left\| \frac{\partial u_N}{\partial y_j}(y) \right\|_{H_0^1} \\ &\leq (\kappa_p + \kappa_{2p}) C_\varphi \sqrt{\lambda_i} \sqrt{\lambda_j} \|b_i\|_\infty \|b_j\|_\infty \left(1 + \sqrt{\frac{a_{max}^N(y)}{a_{min}^N(y)}} ||u_N(y)||_{H_0^1} \right)^2. \end{aligned}$$

The case $|\alpha| = 3, 4$ are treated similary. □

We are now ready to estimate the weak error, i.e. the quantity $\mathbb{E}[\varphi(u_N) - \varphi(u)]$ in L^p-norm. Before stating and proving the estimate on the weak error, we give the basic idea of the proof. To estimate the weak error, we consider the Taylor expansion at order 2 of $\varphi(u_N) - \varphi(u)$ and remark that first order terms and second order terms such that $i \neq j$ are mean-free. In the case where φ is the identity, formally the second order development is :

$$\begin{aligned} u(\omega, x) - u_N(\omega, x) &= u(Y_1(\omega), ..., Y_N(\omega), Y_{N+1}(\omega), ..., x) - u(Y_1(\omega), ..., Y_N(\omega), 0, ..., x) \\ &= \sum_{i>N} \frac{\partial u}{\partial y_i}(Y_1(\omega), ..., Y_N(\omega), 0, ..., x) Y_i(\omega) \\ &+ \frac{1}{2} \sum_{i,j>N} \frac{\partial^2 u}{\partial y_i \partial y_j}(Y_1(\omega), ..., Y_N(\omega), 0, ..., x) Y_i(\omega) Y_j(\omega) + ... \end{aligned}$$

Combining the independence of the Y_i with the fact that the Y_i are mean-free yields that the following terms are mean-free :

$$\begin{aligned} \mathbb{E}\left[\frac{\partial u}{\partial y_i}(Y_1(\omega), ..., Y_N(\omega), 0, ..., x) Y_i(\omega) \right] &= \mathbb{E}\left[\frac{\partial u}{\partial y_i}(Y_1(\omega), ..., Y_N(\omega), 0, ..., x) \right] \mathbb{E}[Y_i(\omega)] \\ &= 0. \end{aligned}$$

Analogously, for $i \neq j$,

$$\begin{aligned} & \mathbb{E}\left[\frac{\partial^2 u}{\partial y_i \partial y_j}(Y_1(\omega), ..., Y_N(\omega), 0, ..., x) Y_i(\omega) Y_j(\omega) \right] \\ = \quad & \mathbb{E}\left[\frac{\partial^2 u}{\partial y_i \partial y_j}(Y_1(\omega), ..., Y_N(\omega), 0, ..., x) \right] \mathbb{E}[Y_i(\omega)] \mathbb{E}[Y_j(\omega)] \\ = \quad & 0. \end{aligned}$$

The proof below shows that indeed the dominant in the error on the expected value is

$$\sum_{i>N} \mathbb{E}\left[\frac{\partial^2 u}{\partial y_i^2}(Y_1(\omega),...,Y_N(\omega),0,...,x)\right]$$

We now give the general and precise result and its proof.

Theorem 2.5.3. *For any $p \leq \infty$ if $d = 1$, $p < \infty$ if $d = 2$, and $p \leq \frac{3}{2}$ if $d = 3$, there exists a constant $C_{2.5.3}(\varphi, p)$ such that for all $N \in \mathbb{N}$, for all $\varphi \in \mathcal{C}^4(\mathbb{R}, \mathbb{R})$ whose derivatives (excluding φ itself) are bounded by a constant C_φ, we have :*

$$\|\mathbb{E}_\omega[\varphi(u_N) - \varphi(u)]\|_{L^p(D)} \leq C_{2.5.3}(\varphi, p) R_N^0$$

Remark 2.5.4. *More generally, the result can be extended to the case where the derivatives of φ are bounded by a polynomial, under extra regularity assumptions on f. This is important since it enables to treat the case of the moments of u. However, this generalization requires additionnal technical difficulties.*

Remark 2.5.5. *The weak error at order N is bounded by R_N^0, whereas the strong error at order N is bounded by $\sqrt{R_N^\alpha}$ for any α as in Assumption 2.3.1. Therefore the weak order is indeed twice the strong order if we can take α as close to zero as we want in Assumption 2.3.1, which is the case in the case of the examples exposed in Section 2.7.*

We give an error analysis in L^p norm. This is of course a weaker result than an error in H^1 norm. A similar result is true with this stronger norm, and in fact it is even true in C^1 norm. However, the proof requires to have a precise estimate of the $C^{1,\beta}$ norm of u_N and its derivatives with respect to the y_i in terms of a_N. This technical estimate follows form a rather long computation.

Proof. Let $M > N$, and $x \in D$, the first order Taylor theorem with integral remainder gives :

$$\begin{aligned}
& \mathbb{E}_\omega[(\varphi(u_M) - \varphi(u_N))(\omega, x)] \\
= \quad & \mathbb{E}_\omega[\varphi(u_M)(Y_1(\omega),...Y_M(\omega), x) - \varphi(u_M)(Y_1(\omega),...Y_N(\omega), 0, ...0, x)] \\
= \quad & \mathbb{E}_\omega[D_y(\varphi \circ u_M)(Y_1(\omega),...Y_N(\omega), 0, ...0, x).(0, ...0, Y_{N+1}(\omega), ..., Y_M(\omega))] \\
+ \quad & \mathbb{E}_\omega\left[\int_0^1 (1-t) D_y^2(\varphi \circ u_M)(Y_1, ...Y_N, tY_{N+1}, ..., tY_M, x).(0, ...0, Y_{N+1}, ..., Y_M)^2 dt\right]
\end{aligned}$$

Since the random variables Y_i are independent, with mean zero and unit variance, the first order term is mean-free :

$$\begin{aligned}
& \mathbb{E}_\omega\left[D_y(\varphi \circ u_M)(Y_1(\omega),...Y_N(\omega), 0, ...0, x).(0, ...0, Y_{N+1}(\omega), ..., Y_M(\omega))\right] \\
= \quad & \mathbb{E}_\omega\left[\sum_{i=N+1}^M \frac{\partial(\varphi \circ u_M)}{\partial y_i}(Y_1(\omega),...Y_N(\omega), 0, ...0, x) Y_i(\omega)\right] \\
= \quad & \sum_{i=N+1}^M \mathbb{E}_\omega\left[\frac{\partial(\varphi \circ u_M)}{\partial y_i}(Y_1(\omega),...Y_N(\omega), 0, ...0, x)\right] \mathbb{E}_\omega[Y_i(\omega)] \\
= \quad & 0
\end{aligned}$$

We now bound the integral remainder term, to begin with we split it into two terms :

$$\begin{aligned}
& \mathbb{E}_\omega\left[(\varphi(u_M) - \varphi(u_N))(\omega, x)\right] \\
= \quad & \sum_{N+1 \leq i,j \leq M} \int_0^1 (1-t) \mathbb{E}_\omega\left[\frac{\partial^2(\varphi \circ u_M)}{\partial y_i \partial y_j}(Y_1, ...Y_N, tY_{N+1}, ..., tY_M, x) Y_i Y_j\right] dt \\
= \quad & \sum_{N+1 \leq i \leq M} \int_0^1 (1-t) \mathbb{E}_\omega\left[\frac{\partial^2(\varphi \circ u_M)}{\partial y_i^2}(Y_1, ...Y_N, tY_{N+1}, ..., tY_M, x) Y_i^2\right] dt \\
+ \quad & \sum_{N+1 \leq i \neq j \leq M} \int_0^1 (1-t) \mathbb{E}_\omega\left[\frac{\partial^2(\varphi \circ u_M)}{\partial y_i \partial y_j}(Y_1, ...Y_N, tY_{N+1}, ..., tY_M, x) Y_i Y_j\right] dt
\end{aligned}$$

First we give an estimate for the first error contribution. Using the bound of the derivatives of $\varphi \circ u_M$ given in Proposition 2.5.2, we get for $N+1 \leq i \leq M$:

$$\begin{aligned}
& \left\| \int_0^1 (1-t)\mathbb{E}_\omega \left[\frac{\partial^2(\varphi \circ u_M)}{\partial y_i^2}(Y_1, ...Y_N, tY_{N+1}, ...tY_M, x) Y_i^2 \right] dt \right\|_{L^p(D)} \\
\leq & \int_0^1 (1-t)\mathbb{E}_\omega \left[\left\| \frac{\partial^2(\varphi \circ u_M)}{\partial y_i^2}(Y_1, ...Y_N, tY_{N+1}, ...tY_M, x) \right\|_{L^p(D)} Y_i^2 \right] dt \\
\leq & C'(2,p) C_\varphi \|b_i\|_\infty^2 \lambda_i \int_0^1 \mathbb{E}_\omega \left[\left(1 + \sqrt{\frac{a_N^{max}}{a_N^{min}}} ||u_M||_{H_0^1} \right)^2 (Y_1, ..., Y_N, tY_{N+1}, ..., tY_M, x) Y_i^2 \right] dt \\
\leq & 2C'(2,p) C_\varphi \|b_i\|_\infty^2 \lambda_i \int_0^1 \mathbb{E}_\omega \left[\left(1 + \frac{a_N^{max}}{a_N^{min}} ||u_M||_{H_0^1}^2 \right) (Y_1, ..., Y_N, tY_{N+1}, ..., tY_M) Y_i^2 \right] dt.
\end{aligned}$$

We define, for $t \in [0,1]$, $\gamma_t \in [0,1]^M$ by $\gamma_t(i) = 1$ for $i \leq N$ and $\gamma_t(i) = t$ for $i > N$, then by Hölder inequality and Proposition 2.3.10 and Lemma 2.4.3 we have :

$$\begin{aligned}
& \mathbb{E}_\omega \left[\frac{a_M^{max}}{a_M^{min}} \|u_M\|_{H_0^1}^2 (Y_1, ..., Y_N, tY_{N+1}, ..., tY_M, x) Y_i^2 \right] \\
\leq & \|a_{\gamma_t, M}^{max}\|_{L^4(\Omega)} \left\| \frac{1}{a_{\gamma_t, M}^{min}} \right\|_{L^4(\Omega)} \|u_M^{\gamma_t}\|_{L^8(\Omega, H_0^1)}^2 \|Y_i\|_{L^8(\Omega)}^2 \\
\leq & D_4^2 G_8^2 m_8^2.
\end{aligned}$$

Where m_8 is the moment of order 8 of a gaussian with mean zero and unit variance.
We obtain finally the following bound for the first term of the error contribution.

$$\begin{aligned}
& \left\| \sum_{N+1 \leq i \leq M} \int_0^1 (1-t)\mathbb{E}_\omega \left[\frac{\partial^2 u_M}{\partial y_i^2}(Y_1, ..., Y_N, tY_{N+1}, ..., tY_M, x) Y_i^2 \right] dt \right\|_{L^p(D)} \\
\leq & 2C'(2,p) C_\varphi (m_4 + D_4^2 G_8^2 m_4) \sum_{N+1 \leq i \leq M} \lambda_i \|b_i\|_\infty^2 \\
\leq & C_\varphi k_1(p) R_N^0.
\end{aligned}$$

Where $k_1(p) = 2C'(2,p)C^2(m_4 + D_4^2 G_8^2 m_8^2)$. Next we give an estimate for the second term of the error contribution, by using once again the independence of the random variables Y_i, for $N+1 \leq i < j \leq M$ we

get :

$$\begin{aligned}
&\int_0^1 (1-t)\mathbb{E}\left[\frac{\partial^2(\varphi\circ u_M)}{\partial y_i\partial y_j}(X_{i,j}^{t,1,1},x)Y_iY_j\right]dt\\
=&\int_0^1 (1-t)\mathbb{E}\left[\frac{\partial^2(\varphi\circ u_M)}{\partial y_i\partial y_j}(X_{i,j}^{t,1,1},x)Y_iY_j\right]dt\\
-&\int_0^1 (1-t)\mathbb{E}\left[\frac{\partial^2(\varphi\circ u_M)}{\partial y_i\partial y_j}(X_{i,j}^{t,0,1},x)Y_iY_j\right]dt\\
=&\mathbb{E}\left[\iint_{[0,1]^2}(1-t)(1-u)t\frac{\partial^3(\varphi\circ u_M)}{\partial y_i^2\partial y_j}(X_{i,j}^{t,u,1},x)Y_i^2Y_jdtdu\right]\\
=&\mathbb{E}\left[\iint_{[0,1]^2}(1-t)(1-u)t\frac{\partial^3(\varphi\circ u_M)}{\partial y_i^2\partial y_j}(X_{i,j}^{t,u,1},x)Y_i^2Y_jdtdu\right]\\
-&\mathbb{E}\left[\iint_{[0,1]^2}(1-t)(1-u)t\frac{\partial^3(\varphi\circ u_M)}{\partial y_i^2\partial y_j}(X_{i,j}^{t,u,0},x)Y_i^2Y_jdtdu\right]\\
=&\mathbb{E}\left[\iiint_{[0,1]^3}(1-t)(1-u)(1-s)t^2\frac{\partial^4(\varphi\circ u_M)}{\partial y_i^2\partial y_j^2}(X_{i,j}^{t,u,s},x)Y_i^2Y_j^2dtduds\right].
\end{aligned}$$

Where the random variables $X_{i,j}^{t,u,s}(\omega)$ are defined by

$$X_{i,j}^{t,u,s}(\omega)=(Y_1,...,Y_N,tY_{N+1},...,tuY_i,...,tsY_j,...,tY_M)(\omega).$$

By Proposition 2.5.2, we have then :

$$\begin{aligned}
&\left\|\mathbb{E}\left[\iiint_{[0,1]^3}(1-t)(1-u)(1-s)t^2\frac{\partial^4(\varphi\circ u_M)}{\partial y_i^2\partial y_j^2}(X_{i,j}^{t,s,u},x)Y_i^2Y_j^2dtduds\right]\right\|_{L^p(D)}\\
\leq&\iiint_{[0,1]^3}(1-t)(1-u)(1-s)t^2\mathbb{E}\left[\left\|\frac{\partial^4(\varphi\circ u_M)}{\partial y_i^2\partial y_j^2}(X_{i,j}^{t,s,u},x)\right\|_{L^p(D)}Y_i^2Y_j^2\right]dtduds\\
\leq&C'(4,p)\|b_i\|_\infty^2\|b_j\|_\infty^2C_\varphi\lambda_i\lambda_j\iiint_{[0,1]^3}\mathbb{E}\left[\left(1+\|u_M\|_{H_0^1}\sqrt{\frac{a_M^{max}}{a_M^{min}}}\right)^4(X_{i,j}^{t,s,u},x)Y_i^2Y_j^2\right]dtduds\\
\leq&2^3C'(4,p)\|b_i\|_\infty^2\|b_j\|_\infty^2C_\varphi\lambda_i\lambda_j\left(m_4+\iiint_{[0,1]^3}\mathbb{E}\left[\left(\|u_M\|_{H_0^1}\sqrt{\frac{a_M^{max}}{a_M^{min}}}\right)^4(X_{i,j}^{t,s,u},x)Y_i^2Y_j^2\right]dtduds\right).
\end{aligned}$$

We define, for $t,s,u\in[0,1]$, $\gamma_{t,s,u}\in[0,1]^M$ by $\gamma_{t,s,u}(n)=1$ for $n\leq N$, $\gamma_{t,s,u}(n)=t$ for $n>N$ such that $n\neq i, n\neq j$, $\gamma_{t,s,u}(i)=tu$, and $\gamma_{t,s,u}(j)=ts$.
Then, Hölder inequality combined with Proposition 2.3.10 and Lemma 2.4.3 yields the following estimate :

$$\begin{aligned}
&\mathbb{E}_\omega\left[\left(\sqrt{\frac{a_M^{max}}{a_M^{min}}}\|u_M\|_{H_0^1}\right)^4(X_{i,j}^{t,s,u},x)Y_i^2Y_j^2\right]\\
\leq&\|a_{\gamma_{t,s,u},M}^{max}\|_{L^{10}(\Omega)}^2\left\|\frac{1}{a_{\gamma_{t,s,u},M}^{min}}\right\|_{L^{10}(\Omega)}^2\|u_M^{\gamma_{t,s,u}}\|_{L^{20}(\Omega,H_0^1)}^4\|Y_i^2\|_{L^5(\Omega)}\|Y_j^2\|_{L^5(\Omega)}\\
\leq&D_{10}^4\|u_M^{\gamma_{t,s,u}}\|_{L^{20}(\Omega,H_0^1)}^4m_{10}^{\frac{2}{5}}\\
\leq&D_{10}^4G_{20}^4m_{10}^{\frac{2}{5}}.
\end{aligned}$$

We have finally the following estimate for the second term of the error contribution :

$$\begin{aligned}
&\left\| \sum_{N+1\le i\neq j\le M} \int_0^1 (1-t)\mathbb{E}\left[\frac{\partial^2 u_M}{\partial y_i \partial y_j}(Y_1,...Y_N,tY_{N+1},...,tY_M,x)Y_iY_j\right]dt \right\|_{L^p(D)} \\
\le\ & 2^7 C'(4,p)C_\varphi(m_4 + D_{10}^4 G_{20}^4 m_{10}^{\frac{2}{5}}) \sum_{N+1\le i\neq j\le M} \lambda_i\lambda_j \|b_i\|_\infty^2 \|b_j\|_\infty^2 \\
\le\ & C_\varphi k_2(p) \sum_{N+1\le i\neq j\le M} \lambda_i\lambda_j \|b_i\|_\infty^2 \|b_j\|_\infty^2.
\end{aligned}$$

Where $k_2(p) = 2^7 C'(4,p)(m_4 + D_{10}^4 G_{20}^4 m_{10}^{\frac{2}{5}})$.
We have finally the following estimate for the total error :

$$\begin{aligned}
&\mathbb{E}_\omega\left[(\varphi(u_M)-\varphi(u_N))(\omega,x)\right] \\
=\ & \sum_{N+1\le i\le M} \int_0^1 (1-t)\mathbb{E}_\omega\left[\frac{\partial^2 u_M}{\partial y_i^2}(Y_1,...Y_N,tY_{N+1},...,tY_M,x)Y_i^2\right]dt \\
+\ & \sum_{N+1\le i\neq j\le M} \int_0^1 (1-t)\mathbb{E}_\omega\left[\frac{\partial^2 u_M}{\partial y_i \partial y_j}(Y_1,...Y_N,tY_{N+1},...,tY_M,x)Y_iY_j\right]dt.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\|\mathbb{E}_\omega[(\varphi(u_M)-\varphi(u_N))(\omega,x)]\|_{L^p(D)} \\
\le\ & C_\varphi k_1 R_N^0 + C_\varphi k_2 \sum_{N+1\le i\neq j\le M} \lambda_i\lambda_j \|b_i\|_\infty^2 \|b_j\|_\infty^2 \\
\le\ & C_\varphi k_1(p) R_N^0 + k_2(p) C_\varphi \left(\sum_{N+1\le i\le M} \lambda_i \|b_i\|_\infty^2\right)^2 \\
\le\ & C_\varphi(k_1(p) + k_2(p)R_0^0)R_N^0.
\end{aligned}$$

We define then $C_{2.5.3}(\varphi,p) = C_\varphi(k_1(p) + k_2(p)R_0^0)$.
We are now ready to conclude, by giving a bound for $\mathbb{E}_\omega[\varphi(u)-\varphi(u_N)]$ in L^p-norm.
Let $N \ge 1$, then for all $M > N$, we have :

$$\begin{aligned}
||\mathbb{E}_\omega[\varphi(u)-\varphi(u_N)]||_{L^p} &\le ||\mathbb{E}_\omega[(\varphi(u)-\varphi(u_M))]||_{L^p} + ||\mathbb{E}_\omega[(\varphi(u_M)-\varphi(u_N)))]||_{L^p} \\
&\le ||\varphi(u)-\varphi(u_M)||_{L^2(\Omega,L^p(D))} + C_{2.5.3}(\varphi,p)R_N^0 \\
&\le C_\varphi||u-u_M||_{L^2(\Omega,L^p(D))} + C_{2.5.3}(\varphi,p)R_N^0.
\end{aligned}$$

Letting $M \to +\infty$, by Proposition 2.4.2 we have :

$$||\mathbb{E}_\omega[\varphi(u)-\varphi(u_N)]||_{L^p} \le C_{2.5.3}(\varphi,p)R_N^0.$$

□

Remark 2.5.6. *Note that the independance of the Y_i is crucial in the proof of* 2.5.3. *This is always the case when one takes the Karhunen-Loève expansion of a gaussian field.*

2.6 An estimate of the total error for the collocation method

In this section, we recall the stochastic collocation method and slightly generalize the proof of the convergence result given in [1] to the case considered here where the assumption of uniform coercivity with respect

to ω is not valid. Since we use many preliminary results of [1], we keep their framework and their notations. As we will see later, it is unfortunately difficult to get a relevant estimate of the dependance with respect to N of the constant appearing by bounding the collocation error.
With the same notations and assumptions as above, we have a regularity result for the solution u_N with respect to y. We introduce a weight $\sigma(y) = \prod_{n=1}^N \sigma_n(y_n)$, where $\sigma_n(y_n) = e^{-\alpha_n|y_n|}$, for any $\alpha \in \mathbb{R}^N$ such that $\alpha_n \geq \sqrt{\lambda_n}\|b_n\|_\infty$ for all $1 \leq n \leq N$, and the functional space

$$\mathcal{C}^0_\sigma(\mathbb{R}^N, V) = \{v : \mathbb{R}^N \to V,\ v \text{ continuous in } y,\ \sup_{y\in\mathbb{R}^N} \|\sigma(y)v(y)\|_V < +\infty\},$$

for any Banach space V. We denote by ρ the density of $Y = (Y_1, ..., Y_N)$, we have then $\rho(y) = \prod_{n=1}^N \rho_n(y) = \frac{1}{(2\pi)^{N/2}} e^{-\sum_{n=1}^N \frac{y_n^2}{2}}$. The following inclusion holds true :

$$\mathcal{C}^0_\sigma(\mathbb{R}^N, H^1_0(D)) \subset L^2_\rho(\mathbb{R}^N, H^1_0(D)),$$

with continuous embedding. More precisely, for any $v \in \mathcal{C}^0_\sigma(\mathbb{R}^N, H^1_0(D))$, we have

$$\|v\|_{L^2_\rho(\mathbb{R}^N, H^1_0(D))} \leq k_\alpha \|v\|_{\mathcal{C}^0_\sigma(\mathbb{R}^N, H^1_0(D))}$$

where $k_\alpha = \prod_{n=1}^N \frac{1}{\sqrt{2\pi}} \int e^{-\frac{y_n^2}{2} + 2\alpha_n|y_n|} dy_n$.

Proposition 2.6.1. *The solution u_N of the equation :*

$$\begin{cases} -\mathrm{div}_x(a_N(y,x)\nabla_x u_N(y,x)) &= f(x) \quad \text{on } D, \\ u_N(y,x) &= 0 \quad \text{on } \partial D, \end{cases}$$

satisfies $u_N \in \mathcal{C}^0_\sigma(\mathbb{R}^N, H^1_0(D))$.

Proof. We recall that $y \mapsto u(y)$ is continuous from $\mathbb{R}^N$ to $H^1_0(D)$. For all $y \in \mathbb{R}^N$, since $\frac{1}{a_N^{min}(y)} \leq e^{\sum_{n=1}^N \sqrt{\lambda_n}\|b_n\|_\infty |y_n|}$, by (2.1) we have :

$$\begin{aligned} \sigma(y)\|u_N(y)\|_{H^1_0(D)} &\leq \sigma(y)\frac{C_D}{a_N^{min}(y)}\|f\|_{L^2(D)} \\ &\leq C_D\|f\|_{L^2(D)} e^{\sum_{n=1}^N (\sqrt{\lambda_n}\|b_n\|_\infty - \alpha_n)|y_n|} \\ &\leq C_D\|f\|_{L^2(D)}. \end{aligned}$$

□

We now give an analyticity result of the solution u_N with respect to y, based on one-dimensional arguments in each direction y_n. We introduce the following notation : for any $1 \leq n \leq N$ and $y \in \mathbb{R}^N$, $y = (y_n, y_n^*)$, where $y_n^* \in \mathbb{R}^{N-1}$. We set $\rho_n^*(y) = \prod_{j\neq n} \rho_j(y)$ and $\sigma_n^*(y) = \prod_{j\neq n} \sigma_j(y)$.

Proposition 2.6.2. *For any $1 \leq n \leq N$, the solution $u_N(y_n, y_n^*, x)$ as a function of y_n, $u : \mathbb{R} \to \mathcal{C}^0_{\sigma_n^*}(\mathbb{R}^{N-1}, H^1_0(D))$ admits an analytic extension $u(z, y_n^*, x)$, $z \in \mathbb{C}$, in the region of the complex plane $\Sigma(\tau) = \{z \in \mathbb{C},\ dist(z, \mathbb{R}) \leq \tau\}$, for $\tau < r_n = \frac{1}{2\|b_n\|_\infty\sqrt{\lambda_n}}$, moreover, for all $z \in \Sigma(\tau)$,*

$$\|\sigma_n(\mathrm{Re}\ z)u_N(z)\|_{\mathcal{C}^0_{\sigma_n^*}(\mathbb{R}^{N-1}, H^1_0(D))} \leq \frac{C_D}{1 - 2\|b_n\|_\infty\sqrt{\lambda_n}\tau_n}\|f\|_{L^2(D)} e^{\alpha_n\tau_n}. \tag{2.5}$$

Proof. For any $y \in \mathbb{R}^N$, $u_N(y)$ satisfies the following variational formulation :

$$\int_D a_N(y,x)\nabla u_N(y,x)\nabla v(x)dx = \int_D f(x)v(x)dx \quad \forall v \in H^1_0(D).$$

Therefore, for every $y \in \mathbb{R}^N$, for any $k \geq 1$, the kth derivative of u_N with respect to y_n satisfies the following variational formulation :

$$\begin{aligned} & \int_D a_N(y,x)\nabla \frac{\partial^k u_N}{\partial y_n^k}(y,x)\nabla v(x)dx \\ = \quad & -\sum_{l=1}^{k} \left(\begin{array}{c} k \\ l \end{array} \right) \int_D \frac{\partial^l a_N}{\partial y_n^l}(y,x)\nabla \frac{\partial^{k-l} u_N}{\partial y_n^{k-l}}(y,x)\nabla v(x)dx \quad \forall v \in H_0^1(D). \end{aligned}$$

Since

$$\left| \frac{\partial^l a_N}{\partial y_n^l}(y,x) \right| \leq (\sqrt{\lambda_n}\|b_n\|_\infty)^l |a_N(y,x)| \leq (\sqrt{\lambda_n}\|b_n\|_\infty)^l |a_N(y,x)| l!,$$

we obtain the recursive inequalities

$$\left\| \sqrt{a_N(y,x)}\nabla \frac{\partial^k u_N}{\partial y_n^k}(y,x) \right\|_{L^2(D)} \leq \sum_{l=1}^{k} \left(\begin{array}{c} k \\ l \end{array} \right) (\|b_n\|_\infty \sqrt{\lambda_n})^l l! \left\| \sqrt{a_N(y,x)}\nabla \frac{\partial^{k-l} u_N}{\partial y_n^k}(y,x) \right\|_{L^2(D)},$$

$$\frac{\left\| \sqrt{a_N(y,x)}\nabla \frac{\partial^k u_N}{\partial y_n^k}(y,x) \right\|_{L^2(D)}}{k!} \leq \sum_{l=1}^{k} (\|b_n\|_\infty \sqrt{\lambda_n})^l \frac{\| \sqrt{a_N(y,x)}\nabla \frac{\partial^{k-l} u_N}{\partial y_n^{k-l}}(y,x) \|_{L^2(D)}}{(k-l)!}.$$

A recurrence yields :

$$\frac{\| \sqrt{a_N(y,x)}\nabla \frac{\partial^k u_N}{\partial y_n^k}(y,x) \|_{L^2(D)}}{k!} \quad \leq \quad \| \sqrt{a_N(y,x)}\nabla u_N(y,x) \|_{L^2(D)} (2\|b_n\|_\infty \sqrt{\lambda_n})^k.$$

And finally :

$$\frac{\| \frac{\partial^k u_N}{\partial y_n^k}(y,.) \|_{H_0^1(D)}}{k!} \leq \frac{C_D}{\sqrt{a_N^{min}(y)}} \|f\|_{L^2(D)} (2\|b_n\|_\infty \sqrt{\lambda_n})^k.$$

We now define for every $y_n \in \mathbb{R}$ the power series $u_N : \mathbb{C} \to \mathcal{C}^0_{\sigma_n^*}(\mathbb{R}^{N-1}, H_0^1(D))$ as

$$u_N(z, y_n^*, x) = \sum_{k=0}^{+\infty} \frac{(z-y_n)^k}{k!} \frac{\partial^k u_N}{\partial y_n^k}(y_n, y_n^*, x).$$

Since

$$\left\| \frac{|z-y_n|^k}{k!} \frac{\partial^k u_N}{\partial y_n^k}(y_n, y_n^*, x) \right\|_{\mathcal{C}^0_{\sigma_n^*}(\mathbb{R}^{N-1}, H_0^1(D))} \leq C_D \|f\|_{L^2(D)} e^{\frac{\sqrt{\lambda_n}\|b_n\|_\infty |y_n|}{2}} (2\|b_n\|_\infty \sqrt{\lambda_n})^k |z-y_n|^k$$

the radius of convergence of this series is $r_n = \frac{1}{2\|b_n\|_\infty \sqrt{\lambda_n}}$. Moreover, take $\tau_n \in]0, r_n[$, since for all $z \in \mathbb{C}$ such that $|z-y_n| \leq \tau$, $\sigma_n(\mathrm{Re}(z)) \leq e^{\alpha_n \tau} \sigma_n(y_n)$ we have the following estimate :

$$\begin{aligned} \|\sigma_n(\mathrm{Re}(z)) u_N(z)\|_{\mathcal{C}^0_{\sigma_n^*}(\mathbb{R}^{N-1}, H_0^1(D))} \quad & \leq \quad \frac{C_D}{1 - 2\|b_n\|_\infty \sqrt{\lambda_n}\tau_n} \|f\|_{L^2(D)} e^{\left(\frac{\sqrt{\lambda_n}\|b_n\|_\infty}{2} - \alpha_n \right) |y_n|} e^{\alpha_n \tau_n} \\ & \leq \quad \frac{C_D}{1 - 2\|b_n\|_\infty \sqrt{\lambda_n}\tau_n} \|f\|_{L^2(D)} e^{\alpha_n \tau_n}. \end{aligned}$$

Hence, by a continuation argument, the function $u_N(y)$ can be extended analytically on the whole region $\Sigma(\tau)$ and estimate (2.5) follows. □

We recall here the stochastic collocation method : we seek a numerical approximation to the exact solution u_N of the equation

$$\left\{ \begin{array}{rll} -\mathrm{div}_x(a_N(y,x)\nabla_x u_N(y,x)) & = f(x) & \text{on } D, \\ u_N(y,x) & = 0 & \text{on } \partial D, \end{array} \right.$$

in a finite dimensional subspace $V_{p,h}$ based on a tensor product, $V_{p,h} = R_p(\mathbb{R}^N) \otimes H_h(D)$, where the following hold :

- $H_h(D) \subset H_0^1(D)$ is a standard finite element space, which contains continuous piecewise polynomials defined on regular triangulations $\mathcal{T}_h$ that have a maximum mesh spacing parameter $h > 0$.
- $R_p(\mathbb{R}^N) \subset L^2_\rho(\mathbb{R}^N)$ is the span of tensor product polynomials with degree at most $p = (p_1, ..., p_N)$ i.e., $R_p(\mathbb{R}^N) = \otimes_{n=1}^N R_{p_n}(\mathbb{R})$, with

$$R_{p_n}(\mathbb{R}) = \text{span}(y_n^m, m = 0, ..., p_n), \quad n = 1, ..., N.$$

We first introduce the semidiscrete approximation $u_N^h : \mathbb{R} \to H_h(D)$, obtained by projecting (2.6) onto the subspace $H_h(D)$, for each $y \in \mathbb{R}$, i.e.

$$\int_D a_N(y,x)\nabla u_N^h(y,x)\nabla v(x)dx = \int_D f(x)v(x)dx, \quad \forall v \in H_h(D). \tag{2.6}$$

The next step consists in collocating (2.6) on the zeros of orthogonal polynomials and building the discrete solution $u_N^{h,p} \in R_p(\mathbb{R}^N) \otimes H_h(D)$ by interpolating in y the collocated solutions. For each dimension $n = 1, ..., N$, let y_{n,k_n}, $1 \le k_n \le p_n + 1$, be the $p_n + 1$ roots of the Hermite polynomial q_{p_n+1} of degree $p_n + 1$, which then satisfies $\int_{\mathbb{R}} q_{p_n+1}(y)v(y)\rho(y)dy = 0$ for all $v \in R_{p_n}(\mathbb{R})$. To any vector of indexes $[k_1, ..., k_N]$ we associate the global index

$$k = k_1 + p_1(k_2 - 1) + p_1p_2(k_3 - 1) + ...$$

and we denote by y_k the point $y_k = [y_{1,k_1}, y_{2,k_2}, ..., y_{N,k_N}] \in \mathbb{R}^N$. We also introduce, for each $n = 1, 2, ..., N$ the Lagrange basis $\{l_{n,j}\}_{j=1}^{p_n+1}$ of the space $R_{p_n}(\mathbb{R})$,

$$l_{n,j} \in R_{p_n}(\mathbb{R}), \quad l_{n,j}(y_{n,k}) = \delta_{jk}, \quad j, k = 1, ..., p_n + 1,$$

where δ_{jk} is the Kronecker symbol, and we set $l_k(y) = \prod_{n=1}^N l_{n,k_n}(y_n)$. The points y_k are then the nodes of the Gaussian quadrature formula associated to the weight ρ. Hence, the final approximation is given by

$$u_N^{h,p}(y,x) = \sum_{k=1}^{N_p} u_N^h(y_k, x)l_k(y),$$

where $u_n^h(y_k, x)$ is the solution of problem (2.6) for $y = y_k$.
Equivalently, if we introduce the Lagrange interpolant operator

$$\mathcal{L}_p : \mathcal{C}^0(\mathbb{R}^N, H_0^1(D)) \to R_p(\mathbb{R}^N) \otimes H_0^1(D),$$

such that

$$\mathcal{L}_p v(y) = \sum_{n=1}^{N} v(y_k)l_k(y), \quad \forall v \in \mathcal{C}^0(\mathbb{R}^N, H_0^1(D)),$$

then we have simply $u_N^{h,p} = \mathcal{L}_p u^h$.
Finally, for any continuous function $g : \mathbb{R}^N \to \mathbb{R}$ we introduce the Gauss quadrature formula $E_\rho^p[g]$ approximating the integral $\int_{\mathbb{R}^N} g(y)\rho(y)dy$ as

$$E_\rho^p[g] = \sum_{k=1}^{N_p} \omega_k g(y_k), \quad \omega_k = \prod_{n=1}^{N} \omega_{k_n}, \quad \omega_{k,n} = \int_{\mathbb{R}} l_{k_n}^2(y)\rho_n(y)dy.$$

This can be used to approximate the law of u_N, i.e. we approximate the expected values $\mathbb{E}[\varphi(u_N(Y(\omega), x))]$ as $\mathbb{E}_\rho^p[\varphi(u_N^h)]$, for some function φ.

Our aim is to give an a priori estimate for the total error $\varepsilon = u_N - u_N^{h,p}$ in the natural norm $L^2_\rho(\mathbb{R}^N, H_0^1(D))$. This total error naturally splits into $\varepsilon = (u_N - u_N^h) + (u_N^h - u_N^{h,p})$.
The first term is a term of space discretization error and can be estimated easily, indeed, for all $y \in \mathbb{R}^N$,

the function $u_N^h(y)$ is the orthogonal projection of $u_N(y)$ onto the subspace $H_h(D)$ with respect to the inner product $(u,v) \mapsto \int_D a_N(y,x)\nabla u(x)\nabla v(x)dx$. Therefore, for all $y \in \mathbb{R}^N$,

$$\begin{aligned}\|(u_N - u_N^h)(y)\|_{H_0^1(D)} &\leq \sqrt{\frac{1}{a_N^{min}(y)}} \inf_{v\in H_h(D)} \left(\int_D a_N(y,x)|\nabla(u_N(y)-v)|^2 dx\right)^{\frac{1}{2}} \\ &\leq \sqrt{\frac{a_N^{max}(y)}{a_N^{min}(y)}} \inf_{v\in H_h(D)} \|u_N(y)-v\|_{H_0^1(D)}.\end{aligned}$$

We can finally conclude for the first term, thanks to the standard approximation estimate for the finite element space $H_h(D)$, there exists a constant C_{fe} such that for any $v \in H^2(D)$ and $h > 0$

$$\min_{w\in H_h(D)} \|v-w\|_{H_0^1(D)} \leq C_{fe} h \|v\|_{H^2(D)}.$$

Since $a_N(y)$ is smooth for any $y \in \mathbb{R}^N$, a precised form of the classical elliptic regularity yields that $u_N(y) \in H^2(D)$ for all $y \in \mathbb{R}^N$, with :

$$\begin{aligned}\|u_N(y)\|_{H^2(D)} &\leq k\frac{\|f\|_{L^2(D)}}{a_N^{min}(y)}\left(1+\frac{a_N^{max}(y)}{a_N^{min}(y)}\right)\left(1+\frac{a_N^{max}(y)}{a_N^{min}(y)}+\frac{\|\nabla a_N(y)\|_{L^\infty(D)}}{a_N^{min}(y)}\right) \\ &\leq k\frac{\|f\|_{L^2(D)}}{a_N^{min}(y)}\left(1+\frac{a_N^{max}(y)}{a_N^{min}(y)}\right)^2 (1+\|\nabla g_N(y)\|_{L^\infty(D)})\end{aligned}$$

where k is a constant independent of f, N and y whose value changes. The proof of the precised form of the elliptic regularity result is given in the appendix (section 2.9). We can then bound the spacial discretization error, using Hölder inequality and proposition 2.3.10 :

$$\begin{aligned}&\|(u_N - u_N^h)(y,x)\|_{L^2_\rho(\mathbb{R}^N,H_0^1(D))} \leq \\ &kC_{fe}h\|f\|_{L^2(D)} \left\| \frac{1}{a_N^{min}(y)}\sqrt{\frac{a_N^{max}(y)}{a_N^{min}(y)}}\left(1+\frac{a_N^{max}(y)}{a_N^{min}(y)}\right)^2 (1+\|\nabla g_N(y)\|_{L^\infty(D)})\right\|_{L^2_\rho(\mathbb{R}^N)} \\ &\leq kC_{fe}\|f\|_{L^2(D)} D_8 (1+D_{16})^5 (1+\|\nabla g_N(y)\|_{L^{16}(\mathbb{R}^N,\mathcal{C}^0(D))})h.\end{aligned}$$

Proposition 2.6.3. *There exists a constant k independent of N such that for all $h > 0$*

$$\|u_N - u_N^h\|_{L^2_\rho(\mathbb{R}^N,H_0^1(D))} \leq k\|f\|_{L^2(D)}(1+\|\nabla g_N(y)\|_{L^{16}_\rho(\mathbb{R}^N,\mathcal{C}^0(D))})h.$$

In particular, if the eigenfunctions b_n are two times continuously differentiable, if the following series is convergent

$$\sum_{n\geq 1} \lambda_n \|\nabla b_n\|_\infty^2 < +\infty$$

and if there exists $0 < \theta < 1$ such that the following series is convergent

$$\sum_{n\geq 1} \lambda_n \|\nabla b_n\|_\infty^{2(1-\theta)} \|D^2 b_n\|_\infty^{2\theta} < +\infty,$$

then there exists a constant k' independent of N such that for all $h > 0$

$$\|u_N - u_N^h\|_{L^2_\rho(\mathbb{R}^N,H_0^1(D))} \leq k'\|f\|_{L^2(D)}h.$$

Otherwise we have the following general rough bound :

$$\|u_N - u_N^h\|_{L^2_\rho(\mathbb{R}^N,H_0^1(D))} \leq k\|f\|_{L^2(D)} h 2^{\frac{N}{16}} e^{8\sum_{n=1}^N \lambda_n \|\nabla b_n\|_\infty^2}.$$

Proof. The first inequality follows from what precedes. Under the additional conditions on the eigenpairs (λ_n, b_n) we can prove, similary as in proposition 2.3.8 that for any $p > 0$ there exists a constant $\tilde{B}_p$ such that for any $N \in \mathbb{N}$, we have

$$\|\nabla g_N(y)\|_{L^p_\rho(\mathbb{R}^N, \mathcal{C}^0(\bar{D}))} \leq \tilde{B}_p.$$

This bound combined with the previous bound yields a bound for the finite element error independent of N. The last bound, available without additional assumptions follows from the first bound :

$$\begin{aligned}
\|u_N - u_N^h\|_{L^2_\rho(\mathbb{R}^N, H^1_0(D))} &\leq k\|f\|_{L^2(D)} \left\| 1 + \sum_{n=1}^{N} \sqrt{\lambda_n}|y_n| \|\nabla b_n\|_\infty \right\|_{L^{16}_\rho(\mathbb{R}^N)} h \\
&\leq k\|f\|_{L^2(D)} h \left(\int_{\mathbb{R}^N} \prod_{n=1}^{N} (2\pi)^{-1/2} e^{16\sqrt{\lambda_n}|y_n|\|\nabla b_n\|_\infty - \frac{y_n^2}{2}} dy \right)^{1/16} \\
&\leq k\|f\|_{L^2(D)} h 2^{\frac{N}{16}} e^{8\sum_{n=1}^{N} \lambda_n \|\nabla b_n\|_\infty^2}.
\end{aligned}$$

□

Remark 2.6.4. *In the general case, the bound explodes as $N \to +\infty$. In particular in the case of an exponential covariance (see further, section 7.1), the bound explodes, which is coherent with the fact that the trajectories of solution u do not belong to H^2 since a does not belong to $\mathcal{C}^1$. However, in the case of an analytic covariance (see further, section 7.2) we can obtain a bound for the finite element error which is independent of N (see section 7.2 combined with the last point of the previous proposition). In the general case, we can obtain a bound independent of N with a lower finite element error order (namely $1/2 - \varepsilon$ for any $\varepsilon > 0$), see [43].*

The second term $u_N^h - u_N^{h,p}$ is an interpolation error, indeed $u_N^{h,p} = \mathcal{L}_p u_N^h$. First, we recall some known results of approximation theory for functions defined on a one-dimensional domain with values in a Banach space denoted by V. We recall that the mono-dimensional weights ρ_1 and σ_1 are defined on $\mathbb{R}$ by $\rho_1(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$ and $\sigma_1(y) = e^{-\alpha|y|}$ for some $\alpha > 0$, and consider a mono-dimensional weight ν such that there exists a constant $a > 0$ with $\nu(y) \geq a e^{-\frac{y^2}{4}}$ The proof of the following three propositions can be found in [1] and in the references therein.

Proposition 2.6.5. *The operator $\mathcal{L}_p : \mathcal{C}^0_\nu(\mathbb{R}, V) \to L^2_{\rho_1}(\mathbb{R}, V)$ is continuous. The norm of this linear continuous operator will be denoted by C_1.*

The following proposition, which is a consequence of a result from Uspensky [68], relates the approximation error $v - \mathcal{L}_p v$ in the L^2_ρ-norm with the best approximation error in the weighted $\mathcal{C}^0_{\sigma_1}$-norm.

Proposition 2.6.6. *For every function $v \in \mathcal{C}^0_\nu(\mathbb{R}, V)$ the interpolation error satisfies*

$$\|v - \mathcal{L}_p v\|_{L^2_{\rho_1}(\mathbb{R}, V)} \leq C_2 \inf_{w \in R_p(\mathbb{R}) \otimes V} \|v - w\|_{\mathcal{C}^0_\nu(\mathbb{R}, V)}$$

with a constant C_2 independent of p.

We now analyze the best approximation error for a function $v : \mathbb{R} \to V$ which admits an analytic extension in the complex plane, in the region $\Sigma(\tau) = \{z \in \mathbb{C}, \quad \text{dist}(z, \mathbb{R}) \leq \tau\}$ for some $\tau > 0$. We still denote the extension by v ; in this case, τ represents the distance between $\mathbb{R}$ and the nearest singularity of $v(z)$ in the complex plane. The following result is a consequence of a result from Hille [41].

Proposition 2.6.7. *Let v be a function in $\mathcal{C}^0_{\sigma_1}(\mathbb{R}, V)$. We assume that v admits an analytic extension in the strip of the complex plane $\Sigma(\tau)$ for some $\tau > 0$, and that*

$$\forall z = (y + iw) \in \Sigma(\tau), \quad \sigma_1(y)\|v(z)\|_V \leq C_v(\tau).$$

Then, for any $\delta > 0$ there exists a constant C, independent of p, and a function $\Theta(p) = O(\sqrt{p})$ such that

$$\min_{w \in R_p(\mathbb{R}) \otimes V} \max_{y \in \mathbb{R}} \left| \|v(y) - w(y)\|_V e^{-\frac{(\delta y)^2}{4}} \right| \leq C\Theta(p) e^{-\tau\delta\sqrt{p}}.$$

We are now ready to prove the following proposition, which gives an estimate of the total interpolation error :

Proposition 2.6.8. *For any $\tau = (\tau_n)_n \in \mathbb{R}^N$ with $\tau_n < r_n = \frac{1}{2\|b_n\|_\infty \sqrt{\lambda_n}}$ for any $1 \leq n \leq N$, there exists a constant $C_{\tau,N}$, independent of h and p such that*

$$\|u_N^h - u_N^{h,p}\|_{L^2_\rho(\mathbb{R}^N, H_0^1(D))} \leq C_{\tau,N} \sum_{n=1}^{N} \sqrt{p_n} e^{-\frac{\tau_n}{\sqrt{2}}\sqrt{p_n}} \|f\|_{L^2(D)}.$$

Remark 2.6.9. *In the case where we choose all the τ_n to be equal (which is possible with $\tau < r = \min_{n \in \mathbb{N}} \frac{1}{2\|b_n\|_\infty \sqrt{\lambda_n}}$ for instance), then we can chose $C_{\tau,N}$ equal to $C_\tau 2^N e^{\frac{N\alpha^2}{2}}$, by using a bound of k_α where we have chosen the α_n to be equal. In the general case, i.e. with a more optimal choice of $(\tau_n)_{1 \leq n \leq N}$, we could try to explicit how $C_{\tau,N}$ depends on N by looking in details at the proof of Propositions 2.6.5 and 2.6.6 and especially Proposition 2.6.7 and the result of Hille from which it follows, but it is out of the scope of this paper.*

Proof. From now on, we are back to the the multi-dimensional problem, we are going to split the total interpolation error into N partial errors linked to one-dimensional interpolation errors.
To begin with, repeating the arguments of Proposition 2.6.2, we obtain that u_N^h has the same regularity with respect to y as the exact solution u_N , i.e. more precisely that u_N^h verifies the same properties as proved for u_N in proposition 2.6.2. We focus on the first direction y_1 and define a one-dimensional interpolation operator
$\mathcal{L}_1 : \mathcal{C}^0_{\sigma_1}(\mathbb{R}, L^2_{\rho_1^*}(\mathbb{R}, H_0^1(D))) \to L^2_{\rho_1}(\mathbb{R}, L^2_{\rho_1^*}(\mathbb{R}, H_0^1(D)))$, by

$$\mathcal{L}_1(y_1, y_1^*, x) = \sum_{k=1}^{p_1+1} v(y_{1,k}, y_1^*, x) l_{1,k}(y_1).$$

Then, the global interpolant $\mathcal{L}_p$ can be written as the composition of two interpolation operators $\mathcal{L}_p = \mathcal{L}_1 \circ \mathcal{L}_p^{(1)}$, where $\mathcal{L}_p^{(1)}$ is the interpolation in all the directions $y_2, y_3, ..., y_N$ except y_1, i.e. $\mathcal{L}_p^{(1)} : \mathcal{C}^0_{\sigma_1^*}(\mathbb{R}^{N-1}, H_0^1(D)) \to L^2_{\rho_1^*}(\mathbb{R}^{N-1}, H_0^1(D))$. We have then

$$\|u_N^h - \mathcal{L}_p u_N^h\|_{L^2_\rho(\mathbb{R}^N, H_0^1)} \leq \|u_N^h - \mathcal{L}_1 u_N^h\|_{L^2_\rho(\mathbb{R}^N, H_0^1)} + \|\mathcal{L}_1(u_N^h - \mathcal{L}_p^{(1)} u_N^h)\|_{L^2_\rho(\mathbb{R}^n, H_0^1)}.$$

Let us bound the first term. We think of u_N^h as a function of y_1 with values in a Banach space V, $u_N^h \in L^2_{\rho_1}(\mathbb{R}, V)$, where $V = L^2_{\rho_1^*}(\mathbb{R}^{N-1}, H_0^1(D))$. The following inclusions hold true :

$$\mathcal{C}^0_{\sigma_1}(\mathbb{R}, V) \subset \mathcal{C}^0_{G_1}(\mathbb{R}, V) \subset L^2_{\rho_1}(\mathbb{R}, V)$$

with $G_1(y_1) = e^{-\frac{y_1^2}{8}}$. Since $G_1 \geq a e^{-\frac{y^2}{4}}$, for some $a > 0$, Proposition 2.6.6 yields the following estimate

$$\|u_N^h - \mathcal{L}_1 u_N^h\|_{L^2_\rho(\mathbb{R}^n, H_0^1(D))} \leq C_2 \inf_{w \in R_{p_1}(\mathbb{R}) \otimes V} \|u_N^h - w\|_{\mathcal{C}^0_{G_1}(\mathbb{R}, V)}.$$

To bound the best approximation error in $\mathcal{C}^0_{G_1}(\mathbb{R}, V)$, we employ the fact that $u_N^h \in \mathcal{C}^0_{\sigma_1}(\mathbb{R}, V)$ and the analyticity result of Proposition 2.6.2, which yields the analitycity of u_N^h as a function of $\mathbb{R}$ into V on $\Sigma(\tau_1)$ for every $\tau_1 < r_1$ and the following bound, for all $z \in \Sigma(\tau_1)$:

$$\sigma_1(z) \|u_N^h(z)\|_{\mathcal{C}^0_{\sigma_1^*}(\mathbb{R}^{N-1}, H_0^1(D))} \leq \frac{C_D}{1 - 2\|b_1\|_\infty \sqrt{\lambda_1} \tau_1} \|f\|_{L^2(D)} e^{\alpha_1 \tau_1},$$

which implies, thanks to the continuous embedding of $\mathcal{C}^0_{\sigma_1^*}(\mathbb{R}^{N-1}, H^1_0(D))$ into $L^2_{\rho_1^*}(\mathbb{R}^{N-1}, H^1_0(D))$, that

$$\sigma_1(\mathrm{z})\|u_N^h(z)\|_{L^2_{\rho_1^*}(\mathbb{R}^{N-1},H^1_0(D))} \leq k_{\alpha_1^*}\frac{C_D}{1-2\|b_1\|_\infty\sqrt{\lambda_1}\tau_1}\|f\|_{L^2(D)}e^{\alpha_1\tau_1}.$$

Recalling that we have fixed $V = L^2_{\sigma_1^*}(\mathbb{R}^{N-1}, H^1_0(D))$ in this proof, we are now ready to conclude with the first term by applying proposition 2.6.7 to $u_N^h \in \mathcal{C}^0_{\sigma_1}(\mathbb{R}, V)$, for $\tau_1 < r_1$ and $\delta_1 = \frac{1}{\sqrt{2}}$, which gives the following bound for the best approximation error :

$$\inf_{w\in R_{p_1}(\mathbb{R})\otimes V}\|u_N^h - w\|_{\mathcal{C}^0_{G_1}(\mathbb{R},V)} \leq C\Theta(p_1)e^{-\frac{\tau}{\sqrt{2}}\sqrt{p_1}}.$$

It gives then the following bound for the first term :

$$\|u_N^h - \mathcal{L}_1 u_N^h\|_{L^2_\rho(\mathbb{R}^n,H^1_0(D))} \leq C_2C\Theta(p_1)e^{-\frac{\tau}{\sqrt{2}}\sqrt{p_1}}.$$

Let us bound the second term. By Proposition 2.6.5 applied with $\nu = \sigma_1$, we bound the second term :

$$\|\mathcal{L}_1(u_N^h - \mathcal{L}_p^{(1)}u_N^h)\|_{L^2_\rho(\mathbb{R}^n,H^1_0(D))} \leq C_1\|u_N^h - \mathcal{L}_p^{(1)}u_N^h\|_{\mathcal{C}^0_{\sigma_1(\mathbb{R},V)}}.$$

The term on the right-hand size is again an interpolation error. Thus we have to bound the interpolation error in all the other $N-1$ directions, uniformly with respect to y_1 (in the weighted norm $\mathcal{C}^0_{\sigma_1}$). We can proceed iteratively, defining an interpolation L^2, bounding the resulting error in the direction y_2, and so on. □

The result and proof for the collocation error is very similar to [1], except that the non-uniform boundedness and coercivity have to be taken into account using the weight σ. Moreover, we have provided a bound for the finite element error making explicit the dependance on N.

2.7 Examples

In this section, we give examples of covariance function $cov[g]$ and the strong and weak convergence results corresponding.

2.7.1 The exponential kernel case on a box

In this subsection, we consider the case where the covariance kernel k is exponential, i.e. $k(x) = \sigma^2 e^{-\frac{x}{\ell}}$ and D is a box, the length ℓ is called the correlation length and the norm choosen on D is $\|x-y\|_1 = \sum_{i=1}^d |x_i - y_i|$. In this case the eigenvalues and the eigenfunctions can be found analytically, we first show that the assumptions done in section 1 are fulfilled and then give the convergence rate for the strong and weak convergence of u_N to u. We first treat the mono-dimensional case, for convenience we suppose that the domain D is $(0,1)$. We have then analytic expressions for the eigenvalues and eigenfunctions, the proof can be found in [76]. We consider the characteristic equation

$$(\ell^2 w^2 - 1)\sin(w) = 2\ell w\cos(w)$$

and denote by $(w_n)_{n\geq 1}$ the sequence of its positive roots sorted in an increasing number, then the eigenvalues of the Karhunen-Loève development can be expressed as $\lambda_n = \frac{2\ell\sigma^2}{\ell^2 w_n^2+1}$ and the eigenfunctions as $b_n(x) = \alpha_n(\sin(w_n x) + \ell w_n\cos(w_n x))$ where $\alpha_n = \frac{1}{\sqrt{(\ell^2 w_n^2+1)/2+\ell}}$. The sequence of the roots (w_n) satisfies $w_n \underset{n\to+\infty}{\sim} n\pi$, which yields the following equivalents for the eigenvalues and eigenfunctions :

$$\lambda_n \underset{n\to+\infty}{\sim} \frac{2\sigma^2}{\ell\pi^2 n^2}$$

$$b_n(x) \underset{n\to+\infty}{\sim} \sqrt{2}\cos(w_n x).$$

We can now show that such a covariance kernel fulfills the Assumptions 2.3.1 and 2.3.6 : $k \in \mathcal{C}^{0,1}(\mathbb{R})$, the eigenfunctions b_n are continuously differentiable, and we have $\|b_n\|_\infty \leq 2\sqrt{2}$, $\|b_n'\|_\infty \leq 2\sqrt{2}(n+1/2)\pi$ for every $n \geq 0$ and for any $0 \leq \alpha < 1/2$, the series $\sum_{n\geq 1} \lambda_n \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha}$ is convergent. Moreover, for any $0 < \alpha < 1/2$ we have

$$R_N^\alpha = \sum_{n>N} \lambda_n \|b_n\|_\infty^{2(1-\alpha)} \|\nabla b_n\|_\infty^{2\alpha} \leq \frac{2\sigma^2}{\ell\pi^2} \sum_{n>N} n^{2\alpha-2} \leq \frac{2\sigma^2}{\ell\pi^2(1-2\alpha)} N^{2\alpha-1}.$$

We can then deduce that Assumption 2.3.5 is fulfilled for $p_0 > \frac{1}{1-2\alpha}$. We can therefore apply the results of sections 3 and 4. The application of Theorem 2.4.2 gives the following strong convergence result :

Proposition 2.7.1. *For every $p > 0$ and $0 < \alpha < 1/2$, u_N converges to u almost surely and there exists a constant $C_{2.7.1}(\alpha, p)$ such that for any $N \in \mathbb{N}$,*

$$\|u - u_N\|_{L^p(\Omega, H_0^1(D))} \leq C_{2.7.1}(\alpha, p) N^{\frac{2\alpha-1}{2}},$$

where $C_{2.7.1}(\alpha, p) = \frac{2\sigma^2}{\ell\pi^2(1-2\alpha)} F_{\alpha,p}$.

Proof. □

On the other hand, Theorem 2.5.3 yields the following weak convergence result :

Proposition 2.7.2. *For any $p \leq \infty$ and $\varphi \in \mathcal{C}^4(\mathbb{R}, \mathbb{R})$ whose derivatives are bounded, then there exists a constant $C(\varphi, p)$ such that for any $N \in \mathbb{N}$ we have :*

$$\|\mathbb{E}[\varphi(u_N) - \varphi(u)]\|_{L^p(D)} \leq C_{2.7.2}(\varphi, p) \frac{1}{N},$$

where $C_{2.7.2}(\varphi, p) = C_{2.5.3}(\varphi, p) \frac{2\sigma^2}{\ell\pi^2}$.

We now treat the case where the spatial dimension is $d = 2$, the following result can be extended for any dimension d. We choose for the sake of simplicity $D = (0,1)^2$, but it can be immediately generalized to the case where D is a box. We denote by $(\mu_n)_{n\geq 1}$ the sequence of the eigenvalues sorted in a decreasing order and by $(c_n)_{n\geq 1}$ the corresponding eigenfunctions. Since the b_n are distinct, for any $n \geq 1$, there exists a unique $(i,j) \in (\mathbb{N}^*)^2$ such that $c_n(x) = b_i(x_1)b_j(x_2)$ and we have then $\mu_n = \lambda_i\lambda_j$. Indeed the eigenvectors in dimension 2 are obtained as the tensor product of the mono-dimensional eigenvectors, since we choose the norm $\|.\|_1$. We can then define the following bijective function :

$$\begin{aligned} g \quad : \quad & (\mathbb{N}^*)^2 \to \mathbb{N}^* \\ & (i,j) \mapsto n \text{ such that } c_n(x) = b_i(x_1)b_j(x_2) \end{aligned}$$

First we notice that $\left(n - \frac{1}{2}\right)\pi \leq w_n \leq \left(n + \frac{1}{2}\right)\pi$, therefore, for any $n \geq 1$

$$\lambda_n \leq \frac{2\ell\sigma^2}{\ell^2\left(n - \frac{1}{2}\right)^2\pi^2} \leq \frac{8\sigma^2}{\ell\pi^2}\frac{1}{n^2}$$

and

$$\lambda_n \geq \frac{2\ell\sigma^2}{\ell^2\left(n + \frac{1}{2}\right)^2\pi^2 + 1} \geq \frac{2\ell\sigma^2}{1 + 4\pi^2\ell^2}\frac{1}{n^2}.$$

We recall that the mono-dimensional eigenfunctions b_n are continuously differentiable and that there exists a constant C such that $\|b_n\|_\infty \leq C$ and $\|b_n'\|_\infty \leq Cn$ for all $n \geq 1$. Therefore, for any $(i,j) \in (\mathbb{N}^*)^2$, $c_{g(i,j)}(x) = b_i(x_1)b_j(x_2)$ is continuously differentiable, $\|c_{g(i,j)}\|_\infty \leq \|b_i\|_\infty \|b_j\|_\infty \leq C^2$, and

$$\|\nabla c_{g(i,j)}\|_\infty = \left\| \begin{pmatrix} b_i'(x_1)b_j(x_2) \\ b_i(x_1)b_j'(x_2) \end{pmatrix} \right\|_\infty \leq C^2(i+j) \leq 2C^2 ij.$$

For any $0 \leq \alpha < 1/2$, we have

$$\begin{aligned}\sum_{ij} \lambda_i\lambda_j \|c_{g(i,j)}\|_\infty^{2(1-\alpha)} \|\nabla c_{g(i,j)}\|_\infty^{2\alpha} &\leq 2C^2 \sum_{ij} \lambda_i\lambda_j (ij)^{2\alpha} \\ &\leq 2C^2 \left(\frac{8l\sigma^2}{l^2\pi^2}\right)^2 n^{2(\alpha-1)} d(n).\end{aligned}$$

where $d(n)$ is the number of divisors of n. Therefore, for any ε such that $1 - 2\alpha > \varepsilon > 0$, there exists a constant $k_{\ell,\varepsilon,\sigma}$ such that

$$\sum_{ij} \lambda_i\lambda_j \|c_{g(i,j)}\|_\infty^{2(1-\alpha)} \|\nabla c_{g(i,j)}\|_\infty^{2\alpha} \leq k_{\ell,\sigma,\varepsilon} n^{2(\alpha-1)+\varepsilon}.$$

This proves that Assumption 2.3.1 is fulfilled. We now bound R_N^α to get almost sure convergence and to bound the strong and weak errors.

Take (i,j) such that ij, then if (p,q) is such that $\lambda_p\lambda_q \geq \lambda_i\lambda_j$ we have :

$$\left(\frac{8\ell\sigma^2}{\ell^2\pi^2}\right)^2 \frac{1}{p^2q^2} \geq \lambda_p\lambda_q \geq \lambda_i\lambda_j \geq \left(\frac{2\ell\sigma^2}{1+4\pi^2\ell^2}\right)^2 \frac{1}{n^2}$$

which yields the following bound for $g(i,j)$:

$$\begin{aligned} g(i,j) &\leq Card\left\{(p,q) | \lambda_p\lambda_q \geq \lambda_i\lambda_j\right\} \\ &\leq Card\left\{(p,q) | pq \leq \frac{4(1+4\pi^2\ell^2)}{\ell^2\pi^2} n\right\} \\ &\leq C'_\ell \, n\log(n), \end{aligned}$$

where C'_ℓ is a constant which depends only on ℓ, because

$$\begin{aligned} Card\{(i,j) | ij \leq n\} &\leq n + \frac{n}{2} + \frac{n}{3} + ... + 1 \\ &\leq n(1+\log(n+1)). \end{aligned}$$

For any $\varepsilon > 0$ there exists a constant $C_{\ell,\varepsilon}$ such that for all $i,j \geq 1$

$$g(i,j) \leq C_{\ell,\varepsilon} (ij)^{1+\varepsilon}.$$

Therefore $g(i,j) > N$ implies $ij > p_{N,\varepsilon} = \left(\frac{N}{C_{\ell,\varepsilon}}\right)^{\frac{1}{1+\varepsilon}}$ for any $\varepsilon > 0$.
Take $0 \leq \alpha < 1/2$ and $\varepsilon > 0$ such that $2\alpha + \varepsilon < 1$,

$$\begin{aligned} R_N^\alpha &= \sum_{n>N} \mu_n \|c_n\|_\infty^{2(\alpha-1)} \|\nabla c_n\|^{2\alpha} \\ &\leq \sum_{n>p_{N,\varepsilon}} \sum_{ij} \lambda_i\lambda_j \|c_{g(i,j)}\|_\infty^{2(1-\alpha)} \|\nabla c_{g(i,j)}\|_\infty^{2\alpha} \\ &\leq k_{l,\sigma,\varepsilon} \sum_{n>p_{N,\varepsilon}} n^{2(\alpha-1)+\varepsilon} \\ &\leq k_{l,\sigma,\varepsilon} \frac{(p_{n,\varepsilon})^{2\alpha+\varepsilon-1}}{1-2\alpha-\varepsilon} \\ &\leq \frac{k_{l,\sigma,\varepsilon}}{(1-2\alpha-\varepsilon) C_{\ell,\varepsilon}^{\frac{2\alpha+\varepsilon-1}{1+\varepsilon}}} N^{\frac{2\alpha+\varepsilon-1}{1+\varepsilon}}. \end{aligned}$$

From this bound, we also deduce that Assumption 2.3.5 is fulfilled.

We can then apply Theorem 2.4.2 which gives a strong convergence result, and Proposition 2.3.6

Proposition 2.7.3. *For every $p > 0$ and $0 < \beta < 1/2$, there exists a constant $C_{2.7.3}(\beta, p)$ such that for any $N \in \mathbb{N}$,*

$$\|u - u_N\|_{L^p(\Omega, H_0^1(D))} \leq C_{2.7.3}(\beta, p) N^{\frac{2\beta-1}{2}},$$

and u_N converges to u almost surely.

On the other hand, Theorem 2.5.3 yields the following weak convergence result :

Proposition 2.7.4. *For any $p < \infty$, $\varepsilon > 0$ and $\varphi \in \mathcal{C}^4(\mathbb{R}, \mathbb{R})$ whose derivatives are bounded , there exists a constant $C_{2.7.4}(\varepsilon, \varphi, p)$ such that for any $N \in \mathbb{N}$, we have :*

$$\|\mathbb{E}[\varphi(u_N) - \varphi(u)]\|_{L^p(D)} \leq C_{2.7.4}(\varepsilon, \varphi, p) N^{-1+\varepsilon}.$$

Remark 2.7.5. *This applys only if the domain is a box. However, for a general domain D, we can use the restriction of the Karhunen-Loève expansion of the gaussian field on a bigger domain which is a box. We notice that it is not a Karhuen-Loève expansion anymore.*

2.7.2 The analytic covariance kernel case

We suppose here that the covariance function $cov[g]$ is analytic on $D \times D$, which is the case of a gaussian covariance function $cov[g](x, y) = \sigma^2 e^{-\frac{\|x-y\|^2}{\ell^2 |D|^2}}$ in particular. Then we have the following result from Frauenfelder, Schwab and Todor given in [24] about the eigenvalues decay and about the decay of the derivatives of the eigenfunctions :

Proposition 2.7.6. *There exists two constants $c_1, c_2 > 0$ such that for all $n \geq 1$*

$$\lambda_n \leq c_1 e^{-c_2 n^{1/d}}.$$

For any $s > 0$ there exists a constant c_s such that for any $n \geq 1$,

$$\|b_n\|_\infty \leq c_s |\lambda_n|^{-s} \text{ and } \|\nabla b_n\|_\infty \leq c_s |\lambda_n|^{-s}.$$

The Assumptions 2.3.1 and 2.3.5 are fulfilled, we can then apply the results of sections 2.4 and 2.5.

Proposition 2.7.7. *For any $0 < s < \frac{1}{2}$, and $p > 0$, there exists a constant $C_{2.7.7}(s, p)$ such that for all $N \in \mathbb{N}$*

$$\|u - u_N\|_{L^p(\Omega, H_0^1(D))} \leq C_{2.7.7}(s, p) \sqrt{\sum_{n>N} \lambda_n^{1-2s}},$$

therefore, for any $0 < s < 1/2$ and $p > 0$ there exists a constant $C'_{2.7.7}(s, p, d)$ depending on p, s, d, c_1 and c_2 such that for any $N \in \mathbb{N}$

$$\|u - u_N\|_{L^p(\Omega, H_0^1(D))} \leq C'_{2.7.7}(s, p, d) N^{\frac{d-1}{2d}} e^{-\frac{c_2(1-2s)}{2} N^{1/d}}.$$

We have also almost sure convergence.

Proposition 2.7.8. *For any $p \leq \infty$ if $d = 1$, $p < \infty$ if $d = 2$, and $p \leq \frac{3}{2}$ if $d = 3$ and $0 < s < \frac{1}{2}$, and for all $\varphi \in \mathcal{C}^4(\mathbb{R}, \mathbb{R})$ whose derivatives are bounded there exists a constant $C_{2.7.8}(\varphi, p, s)$ such that for all $N \in \mathbb{N}$, we have :*

$$\|\mathbb{E}[\varphi(u_N) - \varphi(u)]\|_{L^p(D)} \leq C_{2.7.8}(\varphi, p, s) \sum_{n>N} \lambda_n^{1-2s},$$

therefore, for any $0 < s < \frac{1}{2}$, there exists a constant $C'_{2.7.8}(\varphi, p, s)$ depending on d, s, p, c_1 and c_2 such that for all $N \in \mathbb{N}$, for all $\varphi \in \mathcal{C}^4(\mathbb{R}, \mathbb{R})$ whose derivatives are bounded by a constant C_φ

$$\|\mathbb{E}[\varphi(u_N) - \varphi(u)]\|_{L^p(D)} \leq C'_{2.7.8}(\varphi, p, s) N^{\frac{d-1}{d}} e^{-c_2(1-2s)N^{1/d}}.$$

2.8 Conclusions

In this work we have established estimates of the error on the solution of the elliptic partial differential equation, resulting from the approximation of the lognormal random field a through the truncature of the Karhunen-Loève expansion of $\log(a)$. This approximation is indeed the first step of several numerical methods, in particular galerkin stochastic methods and collocation methods. In these methods, since the computational cost increases very fast with the truncature order N, it is crucial to have good estimates of the error commited on the solution u.
We first showed that the strong error decreases like the error on the Karhunen-Loève expansion of $\log(a)$ in the natural $L^2(\Omega \times D)$ norm, i.e. is bounded by the squared root of the remainder of the eigenvalues series. We next showed that this bound can be improved by looking at the weak error, which is a natural quantity of interest since we are interested on the law of the solution. The bound for the weak error is indeed the square of the previous bound for the strong error.
We complete then this work by generalizing the result of [1], which gives an estimate of the collocation error, to the case considered here where the random field a is neither uniformly bounded nor uniformly coercive with respect to ω.
Finally we show that the strong and weak error results apply to two examples which are important on a practical point of view, the case of an exponential covariance and the case of an analytic covariance, which includes the case of a gaussian covariance in particular. We give then explicit bounds for the error in these two cases which are among the most frequently used to model permeability fields in the context of flow computation in porous media.
The analysis of the dependance of the error on the correlation length ℓ and on the multiplicative factor σ in the covariance is the subject of ongoing research.

2.9 Appendix

In this section, we prove a result providing the dependance with respect to the coefficient of the constant in the classical H^2 regularity result for a linear elliptic PDE.

Theorem 2.9.1. *Let D be a $\mathcal{C}^2$ bounded domain of $\mathbb{R}^d$, $f \in L^2(D)$ and $a \in \mathcal{C}^1(\bar{D})$ such that for any $x \in D$, we have $a_{min} \leq a(x) \leq a_{max}$ and $\|\nabla a(x)\| \leq a'_{max}$. We consider then the following elliptic partial differential equation :*

$$\left\{ \begin{array}{rcll} -\mathrm{div}(a(x)\nabla u(x)) & = & f(x) & \text{on } D \\ u(x) & = & 0 & \text{on } \partial D \end{array} \right. \tag{2.7}$$

This equation admits a unique solution $u \in H^1_0(D)$ by Lax-Milgram theorem. We have then $u \in H^2(D)$, with

$$\|u\|_{H^2(D)} \lesssim \frac{\|f\|_{L^2(D)}}{a^3_{min}}(a_{max} + a'_{max})a_{max}.$$

In the previous bound, as in what follows, the constants are chosen independent of a and f, since we are interested in this dependance. In particular, the symbol "$\lesssim$" means bounded by, up to a constant independent of a and f.

All the remainder of this section will be devoted to the proof of this theorem. The following proof follows the proofs of [9], [39], making explicit the dependance on a. We recall that this classical proof is based on the Nirnenberg translation method and consists in three main steps, which correspond to the following three subsections. In the whole proof, we will use the following notations : let $(e_i)_{1 \leq i \leq d}$ be the canonical basis of $\mathbb{R}^d$, then for any function u on D we define $D^i_h u(x) := \frac{u^+(x) - u^-(x)}{h}$, where $u^+(x) = u(x + he_i/2)$ and $u^-(x) = u(x - he_i/2)$. We notice that $(D^i_h)^* = -D^i_h$ and that $D^i_h(uv) = u^+ D^i_h v + v^- D^i_h u$. Moreover, for any $v \in H^1(D)$, $\|D^i_h v\|_{L^2(D)} \leq \left\| \frac{\partial v}{\partial x_i} \right\|_{L^2(D)}$ and conversely, if $v \in L^2(D)$ is such that for any $h > 0$ we have $\|D^i_h v\|_{L^2(D)} \leq C$ then $\frac{\partial v}{\partial x_i} \in L^2(D)$, and $\left\| \frac{\partial v}{\partial x_i} \right\|_{L^2(D)} \leq C$ (Proposition IX.3 in [9]). Besides this, if $v \in \mathcal{C}^1(D)$ with bounded derivatives then $\|D^i_h v\|_\infty \leq \|\nabla v\|_\infty$.

2.9.1 The case $D = \mathbb{R}^d$

Proposition 2.9.2. *We need here to suppose the coefficient a ti be matrix-valued. Let $f \in L^2(\mathbb{R}^d)$ and $a \in \mathcal{C}_b^1(\mathbb{R}^d, S_d(\mathbb{R}))$ such that for any $x, y \in \mathbb{R}^d$, we have $a_{min}\|y\|^2 \leq a(x)y.y$, for any $x \in \mathbb{R}^d$, we have $\|a(x)\|_{M_d(\mathbb{R})} \leq a_{max}$ and $\|Da(x)\|_{\mathcal{L}(\mathbb{R}^d, M_d(\mathbb{R}))} \leq a'_{max}$. If $u \in H^1(\mathbb{R}^d)$ is a solution of*

$$-\mathrm{div}(a(x)\nabla u(x)) = f(x) \text{ on } \mathbb{R}^d.$$

then $u \in H^2(\mathbb{R}^d)$ with

$$|u|_{H^2(\mathbb{R}^d)} \lesssim \frac{1}{a_{min}^2} a_{max}(|u|_{H^1(\mathbb{R}^d)} + \|f\|_{L^2(\mathbb{R}^d)}).$$

Proof. Let $1 \leq i \leq d$. For any $v, w \in H^1(\mathbb{R}^d)$, we introduce the following bilinear form :

$$\begin{aligned} d(v,w) &:= \int a\nabla v \nabla D_h^i w - \int a\nabla (D_h^i)^* v \nabla w dx \qquad (2.8)\\ &= \frac{1}{h}\int_D a\left(\nabla v\nabla w^+ - \nabla v\nabla w^- + \nabla v^+\nabla w - \nabla v^-\nabla w\right)dx \\ &= \frac{1}{h}\int [(a^- - a)\nabla v^- + (a - a^+)\nabla v^+]\nabla w dx. \end{aligned}$$

We have then

$$|d(v,w)| \leq a'_{max}|v|_{H^1(\mathbb{R}^d)}|w|_{H^1(\mathbb{R}^d)}.$$

$$\begin{aligned} a_{min}|D_h^i u|^2_{H^1(\mathbb{R}^d)} &\leq \int a\nabla D_h^i u \nabla D_h^i u dx \\ &= d(D_h^i u, u) + \int_D a\nabla (D_h^i)^* D_h^i u \nabla u \\ &= d(D_h^i u, u) + \int_D f D_h^i D_h^i u \\ &\leq a'_{max}|D_h^i u|_{H^1(\mathbb{R}^d)}|u|_{H^1(\mathbb{R}^d)} + \|f\|_{L^2(\mathbb{R}^d)}|D_h^i u|_{H^1(\mathbb{R}^d)}. \end{aligned}$$

Therefore, for any $h > 0$

$$|D_h^i u|_{H^1(\mathbb{R}^d)} \leq \frac{1}{a_{min}}(a'_{max}|u|_{H^1(\mathbb{R}^d)} + \|f\|_{L^2(\mathbb{R}^d)}),$$

and we deduce that $\frac{\partial u}{\partial x_i} \in H^1(\mathbb{R}^d)$ with

$$\left|\frac{\partial u}{\partial x_i}\right|_{H^1(\mathbb{R}^d)} \leq \frac{1}{a_{min}}(a'_{max}|u|_{H^1(\mathbb{R}^d)} + \|f\|_{L^2(\mathbb{R}^d)}).$$

□

2.9.2 The case $D = \mathbb{R}^d_+$

We define $\mathbb{R}^d_+ = \{y = (y_1, ..., y_d) \in \mathbb{R}^d | y_d > 0\}$.

Proposition 2.9.3. *Let $f \in L^2(\mathbb{R}^d_+)$ and $a \in \mathcal{C}_b^1(\overline{\mathbb{R}^d_+}, S_d(\mathbb{R}))$ such that for any $x \in \mathbb{R}^d_+$ and $y \in \mathbb{R}^d$, we have $a_{min}\|y\|^2 \leq a(x)y.y$, $\|a(x)\|_{M_d(\mathbb{R})} \leq a_{max}$ and $\|Da(x)\|_{\mathcal{L}(\mathbb{R}^d_+, M_d(\mathbb{R}))} \leq a'_{max}$. If $u \in H^1(\mathbb{R}^d_+)$ is a solution of*

$$\left\{ \begin{array}{rcll} -\mathrm{div}(a(x)\nabla u(x)) &=& f(x) & \text{on } \mathbb{R}^d_+ \\ u(x) &=& 0 & \text{on } \partial\mathbb{R}^d_+ \end{array} \right.$$

then $u \in H^2(\mathbb{R}^d_+)$ with

$$|u|_{H^2(\mathbb{R}^d_+)} \lesssim \frac{1}{a_{min}}\left(1 + \frac{a_{max}}{a_{min}}\right)(a'_{max}|u|_{H^1(\mathbb{R}^d_+)} + \|f\|_{L^2(\mathbb{R}^d_+)}).$$

Proof. For $1 \leq i \leq d-1$, since $\mathbb{R}^d_+$ is invariant through a translation by e_i, the same argument as in the previous Proposition shows that, for $1 \leq i \leq d-1$: $\frac{\partial u}{\partial x_i} \in H^1(\mathbb{R}^d_+)$ with

$$\left|\frac{\partial u}{\partial x_i}\right|_{H^1(\mathbb{R}^d_+)} \leq \frac{1}{a_{min}}(a'_{max}|u|_{H^1(\mathbb{R}^d_+)} + \|f\|_{L^2(\mathbb{R}^d_+)}). \tag{2.9}$$

For $1 \leq i,j \leq d$, and $v \in \mathcal{D}(\mathbb{R}^d_+)$ we introduce the following notation :

$$F_{i,j}(v) = \int_{\mathbb{R}^d_+} a_{i,j}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j}dx.$$

We have then, for $(i,j) \neq (d,d)$, and $v \in \mathcal{D}(\mathbb{R}^d_+)$:

$$F_{i,j}(v) \quad = \quad -\int_{\mathbb{R}^d_+} a_{i,j}\frac{\partial^2 u}{\partial x_i \partial x_j}vdx - \int_{\mathbb{R}^d_+}\frac{\partial a_{i,j}}{\partial x_j}\frac{\partial u}{\partial x_i}vdx,$$

hence, for instance if $i \neq d$ (it is similar if $j \neq d$), using (2.9), we have :

$$\begin{aligned} |F_{i,j}(v)| &\leq a_{max}\left|\frac{\partial u}{\partial x_i}\right|_{H^1(\mathbb{R}^d_+)}\|v\|_{L^2(\mathbb{R}^d_+)} + a'_{max}|u|_{H^1(\mathbb{R}^d_+)}\|v\|_{L^2(\mathbb{R}^d_+} \\ &\leq \left(\frac{a_{\max}}{a_{min}}(a'_{max}|u|_{H^1(\mathbb{R}^d_+)} + \|f\|_{L^2(\mathbb{R}^d_+)}) + a'_{max}|u|_{H^1(\mathbb{R}^d_+)}\right)\|v\|_{L^2(\mathbb{R}^d_+)} \end{aligned}$$

Since u solves the equation, we have, for any $v \in H^1_0(\mathbb{R}^d_+)$

$$\sum_{1\leq i,j\leq d} F_{i,j}(v) = \int_{\mathbb{R}^d_+} fvdx.$$

Therefore, for any $v \in \mathcal{D}(\mathbb{R}^d_+)$

$$\begin{aligned} |F_{d,d}(v)| &\leq \sum_{(i,j)\neq(d,d)} |F_{i,j}(v)| + \|v\|_{L^2(\mathbb{R}^d_+)}\|f\|_{L^2(\mathbb{R}^d_+)} \\ &\lesssim \left(1 + \frac{a_{\max}}{a_{min}}\right)(a'_{max}|u|_{H^1(\mathbb{R}^d_+)} + \|f\|_{L^2(\mathbb{R}^d_+)})\|v\|_{L^2(\mathbb{R}^d_+)}, \end{aligned}$$

which implies that $\frac{\partial}{\partial x_d}\left(a_{d,d}\frac{\partial u}{\partial x_d}\right) \in L^2(\mathbb{R}^d_+)$ with

$$\left\|\frac{\partial}{\partial x_d}\left(a_{d,d}\frac{\partial u}{\partial x_d}\right)\right\|_{L^2(\mathbb{R}^d_+)} \lesssim \left(1 + \frac{a_{\max}}{a_{min}}\right)(a'_{max}|u|_{H^1(\mathbb{R}^d_+)} + \|f\|_{L^2(\mathbb{R}^d_+)}).$$

Moreover, by (2.9), for $1 \leq i \leq d-1$, we have

$$\begin{aligned} \left\|\frac{\partial}{\partial x_i}\left(a_{d,d}\frac{\partial u}{\partial x_d}\right)\right\|_{L^2(\mathbb{R}^d_+)} &\lesssim a_{max}\left|\frac{\partial u}{\partial x_i}\right|_{H^1(\mathbb{R}^d_+)} + a'_{max}|u|_{H^1(\mathbb{R}^d_+)} \\ &\lesssim \left(1 + \frac{a_{\max}}{a_{min}}\right)(a'_{max}|u|_{H^1(\mathbb{R}^d_+)} + \|f\|_{L^2(\mathbb{R}^d_+)}). \end{aligned}$$

Finally, $a_{d,d}\frac{\partial u}{\partial x_d} \in H^1(\mathbb{R}^d_+)$, with

$$\left|a_{d,d}\frac{\partial u}{\partial x_d}\right|_{H^1(\mathbb{R}^d_+)} \lesssim \left(1 + \frac{a_{\max}}{a_{min}}\right)(a'_{max}|u|_{H^1(\mathbb{R}^d_+)} + \|f\|_{L^2(\mathbb{R}^d_+)})$$

Since $a_{d,d} \geq a_{min}$, we deduce that $\frac{\partial u}{\partial x_d} \in H^1(\mathbb{R}^d_+)$ with

$$\begin{aligned}\left|\frac{\partial u}{\partial x_d}\right|_{H^1(\mathbb{R}^d_+)} &= \left|\frac{1}{a_{d,d}}\left(a_{d,d}\frac{\partial u}{\partial x_d}\right)\right|_{H^1(\mathbb{R}^d_+)} \\ &\lesssim \frac{1}{a_{min}}\left|a_{d,d}\frac{\partial u}{\partial x_d}\right|_{H^1(\mathbb{R}^d_+)} + \frac{a'_{max}}{a^2_{min}}\left\|a_{d,d}\frac{\partial u}{\partial x_d}\right\|_{L^2(\mathbb{R}^d_+)} \\ &\lesssim \frac{1}{a_{min}}\left(1+\frac{a_{max}}{a_{min}}\right)(a'_{max}|u|_{H^1(\mathbb{R}^d_+)} + \|f\|_{L^2(\mathbb{R}^d_+)}).\end{aligned}$$

This bound, combined with (2.9), yields the result. □

2.9.3 The case D bounded

We now prove the theorem, using the two previous Propositions in two successive steps. We recall that D is supposed to be $\mathcal{C}^2$.
Let $(D_i)_{0\leq i\leq p}$ be a covering of D such that the $(D_i)_{0\leq i\leq p}$ are open and bounded, $\overline{D} \subset \cup_{i=0}^p D_i$, $\cup_{i=1}^p (D_i \cap \partial D) = \partial D$, $\overline{D}_0 \subset D$.
We define $Q = \{(y', y_d) \in \mathbb{R}^d |\ \|y'\| < 1 \text{ and } |y_d| < 1\}$, $Q_+ = Q \cap \mathbb{R}^d_+$ and $Q_0 = \{(y', y_d) \in \mathbb{R}^d |\ \|y'\| < 1 \text{ and } y_d = 0\}$.
For $1 \leq i \leq p$, let α_i be a bijection from D_i to Q such that $\alpha_i \in \mathcal{C}^2(\overline{D}_i)$, $\alpha_i^{-1} \in \mathcal{C}^2(\overline{Q})$, $\alpha_i(D_i \cap D) = Q_+$ and $\alpha_i(D_i \cap \partial D) = Q_0$.

Let $(\chi_i)_{0\leq i\leq p}$ be an associated partition of unity, i.e. for $0 \leq i \leq p$, we have $\chi_i \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}_+)$, $\sum_{i=0}^p \chi_i = 1$ on $\overline{D}$, and for $0 \leq i \leq d$, $\mathrm{supp}(\chi_i)$ is compact with $\mathrm{supp}(\chi_i) \subset D_i$. We denote by u the solution of (2.7), u can then be split into $u = \sum_{i=0}^p u_i$, with $u_i = u\chi_i$. We treat separately u_0 and the u_i, for $1 \leq i \leq p$ using respectively the whole space case and the half-space case.

Lemma 2.9.4. *u_0 belongs to $H^2(D)$, with*

$$\|u_0\|_{H^2(D)} \lesssim \frac{\|f\|_{L^2(D)}}{a^2_{min}}(a_{max} + a'_{max}).$$

Proof. Since $\mathrm{supp}(u_0) \subset D_0$, $u_0 \in H^1_0(D)$ and besides, u_0 is the solution on D of the new equation : $-\mathrm{div}(a\nabla u_0) = g$, where $g \in L^2(D)$ is defined by

$$\begin{aligned} g &:= f\chi_0 + a\nabla u.\nabla\chi_0 + \mathrm{div}(au\nabla\chi_0) \\ &= f\chi_0 - 2a\nabla u.\nabla\chi_0 - u\nabla a.\nabla\chi_0 - au\Delta\chi_0.\end{aligned}$$

We continue u_0 on $\mathbb{R}^d$ by 0 on $\mathbb{R}^d \setminus D$, which yields $u_0 \in H^1(\mathbb{R}^d)$ and g by 0 on $\mathbb{R}^d \setminus D$, which yields $g \in L^2(\mathbb{R}^d)$. Let $\psi \in \mathcal{C}^\infty(\mathbb{R}^d, [0,1])$ such that $\psi = 0$ on D_0 and $\psi = 1$ on $\tilde{D}^c$, where $\tilde{D}$ is an open set such that $\overline{D}_0 \subset \tilde{D}$ and $\overline{\tilde{D}} \subset D$. We define then $\bar{a}$ by

$$\bar{a}(x) = \begin{cases} a(x)(1-\psi(x)) + a_{min}\psi(x) & \text{if } x \in D \\ a_{min}\psi(x) & \text{if } x \in D^c \end{cases}$$

We have then, $\bar{a} \in \mathcal{C}^1(\mathbb{R}^d)$, and for any $x \in \mathbb{R}^d$,

$$a_{min} \leq \bar{a}(x) \leq a(x) \leq a_{max},$$

and

$$\|\nabla\bar{a}(x)\| \lesssim \bar{a}'_{max} := a'_{max} + a_{max}.$$

Using these extensions, we have $-\mathrm{div}(\bar{a}\nabla u_0) = g$ on $\mathbb{R}^d$. Indeed, for any $v \in \mathcal{D}(\mathbb{R}^d)$,

$$\begin{aligned}\int_{R^d} \bar{a}(x)\nabla u_0(x)\nabla v(x) &= \int_{D_0} \bar{a}(x)\nabla u_0(x)\nabla v(x) \\ &= \int_{D_0} a(x)\nabla u_0(x)\nabla v(x) \\ &= \int_D a(x)\nabla u_0(x)\nabla v(x)\end{aligned}$$

We have used that, since supp(u_0) is included is the open bounded set D_0, we have $\nabla u_0 = 0$ on D_0^c and $a = \bar{a}$ on D_0.
We recall that, by Poincaré inequality,

$$\|u\|_{L^2(D)} \leq C_D |u|_{H^1(D)},$$

and classically, using the variational formulation,

$$|u|_{H^1(D)} \leq \frac{1}{a_{min}} \|f\|_{L^2(D)}.$$

Therefore

$$\begin{aligned} |u_0|_{H^1(\mathbb{R}^d)} &\leq |u|_{H^1(D)} \|\chi_0\|_\infty + \|u\|_{L^2(D)} \|\nabla \chi_0\|_\infty \\ &\lesssim \frac{\|f\|_{L^2(D)}}{a_{min}}, \end{aligned}$$

and

$$\begin{aligned} \|g\|_{L^2(\mathbb{R}^d)} &\leq \|f\|_{L^2(D)} \|\chi_0\|_\infty + 2a_{max} |u|_{H^1(D)} \|\nabla \chi_0\|_\infty \\ &+ \|u\|_{L^2(D)} a'_{max} \|\nabla \chi_0\|_\infty + a_{max} \|u\|_{L^2(D)} \|\Delta \chi_0\|_\infty \\ &\lesssim \frac{\|f\|_{L^2(D)}}{a_{min}} (a_{max} + a'_{max}). \end{aligned}$$

We can then apply Proposition 2.9.2 with $\bar{a}$, u_0 and g, which yields that $u_0 \in H^2(\mathbb{R}^d)$ with

$$\begin{aligned} |u_0|_{H^2(D)} &\lesssim \frac{1}{a_{min}} (\bar{a}'_{max} |u_0|_{H^1(\mathbb{R}^d)} + \|g\|_{L^2(D)}) \\ &\lesssim \frac{1}{a_{min}} \left[(a'_{max} + a_{max}) \frac{\|f\|_{L^2(D)}}{a_{min}} + \frac{\|f\|_{L^2(D)}}{a_{min}} (a_{max} + a'_{max}) \right] \\ &\lesssim \frac{\|f\|_{L^2(D)}}{a_{min}^2} (a_{max} + a'_{max}), \end{aligned}$$

where the constants depend on χ_0, ψ and on the constant of Poincaré inequality C_D, i.e. on D, but neither on a or f. □

We now treat the cases of the $(u_i)_{1 \leq i \leq p}$.

Lemma 2.9.5. *For $1 \leq i \leq p$, $u_i \in H^2(D)$, with*

$$\|u_i\|_{H^2(D)} \lesssim \frac{\|f\|_{L^2(D)}}{a_{min}^3} (a'_{max} + a_{max}) a_{max}.$$

Proof. Similarly to the proof of the previous Lemma, we have that
$u_i \in H_0^1(D \cap D_i)$ solves $-\mathrm{div}(a \nabla u_i) = g_i$ on $D \cap D_i$, where $g_i \in L^2(D \cap D_i)$ is defined by

$$\begin{aligned} g_i &:= f\chi_i + a \nabla u . \nabla \chi_i + \mathrm{div}(au \nabla \chi_i) \\ &= f\chi_i - 2a \nabla u . \nabla \chi_i - u \nabla a . \nabla \chi_i - au \Delta \chi_i. \end{aligned}$$

Analogously to the case of g, we have

$$\|g_i\|_{L^2(D \cap D_i)} \lesssim \frac{\|f\|_{L^2(D)}}{a_{min}} (a_{max} + a'_{max}).$$

We now define, for $y \in Q_+$, $v_i(y) = u_i(\alpha_i^{-1}(y))$. Let $\varphi \in H_0^1(Q_+)$, we define for $x \in D_i \cap D$, $v(x) = \varphi(\alpha_i(x))$. We have then $v \in H_0^1(D_i \cap D)$ with $\nabla v(x) = D\alpha_i^t(x) \nabla \varphi(\alpha_i(x))$ and $v_i \in H_0^1(Q_+)$ with $\nabla v_i(y) =$

$(D\alpha_i^t)^{-1}(\alpha_i^{-1}(y))\nabla u_i(\alpha_i^{-1}(y))$.

$$\begin{aligned}
&\int_{Q_+} g_i(\alpha_i^{-1}(y))\varphi(y)|\det(D\alpha_i^{-1}(y))|dy \\
&= \int_{D_i\cap D} g_i(x)v(x)dx \\
&= \int_{D_i\cap D} a(x)\nabla u_i(x).\nabla v(x)dx \\
&= \int_{Q_+} a(\alpha_i^{-1}(y))(\nabla u_i)(\alpha_i^{-1}(y)).(\nabla v)(\alpha_i^{-1}(y))|\det(D\alpha_i^{-1}(y))|dy \\
&= \int_{Q_+} a(\alpha_i^{-1}(y))(D\alpha_i^t)(\alpha_i^{-1}(y))\nabla v_i(y).(D\alpha_i^t)(\alpha_i^{-1}(y))\nabla\varphi(y)|\det(D\alpha_i^{-1}(y))|dy \\
&= \int_{Q_+} A_i(y)\nabla v_i(y).\nabla\varphi(y)dy,
\end{aligned}$$

with the definition $A_i(y) = a(\alpha_i^{-1}(y))|\det(D\alpha_i^{-1}(y))|(D\alpha_i D\alpha_i^t)(\alpha_i^{-1}(y)) \in S_d(\mathbb{R})$. Denoting $F_i(y) = g_i(\alpha_i^{-1}(y))|\det(D\alpha_i^{-1}($
which gives $F_i \in L^2(Q_+)$. We get finally that $v_i \in H_0^1(Q_+)$ solves : for any $\varphi \in H_0^1(Q_+)$,

$$\int_{Q_+} A_i(y)\nabla v_i(y).\nabla\varphi(y)dy = \int_{Q_+} F_i(y)\varphi(y).$$

In order to apply Lemma 2.9.3, we first check that A_i is $\mathcal{C}^1$ and coercive, and then define extensions of A_i and F_i to $\mathbb{R}_+^d$. For any $y \in Q_+$, $A_i(y) = a(\alpha_i^{-1}(y))|\det(D\alpha_i^{-1}(y))|(D\alpha_i D\alpha_i^t)(\alpha_i^{-1}(y))$, therefore $A_i \in \mathcal{C}_b^1(\overline{Q}_+, M_d)$. Moreover, for any $y \in Q_+$, $z \in \mathbb{R}^d$, recalling that α_i is a $\mathcal{C}^2$ diffeomorphism from $D_i \cap D$ to Q_+, with $\alpha_i^{-1} \in \mathcal{C}^2(\overline{Q}_+)$, we have the coercivity property :

$$\begin{aligned}
A_i(y)z.z &= a(\alpha_i^{-1}(y))|\det(D\alpha_i^{-1}(y))|\ \|(D\alpha_i^t)(\alpha_i^{-1}(y))z\|_2^2 \\
&\geq a_{min}|\det(D\alpha_i^{-1}(y))|\ \|(D\alpha_i^t)^{-1}(\alpha_i^{-1}(y))\|_2^{-2}\|z\|_2^2 \\
&\geq a_{min}\min_{y\in\overline{Q}_+}\{|\det(D\alpha_i^{-1}(y))|\ \|(D\alpha_i^t)^{-1}(\alpha_i^{-1}(y))\|_2^{-2}\}\|z\|_2^2 \\
&\gtrsim a_{min}\|z\|_2^2,
\end{aligned}$$

the boundedness property :

$$\begin{aligned}
\|A_i(y)\|_{M_d} &\leq a_{max}\max_{y\in\overline{Q}_+}\{|\det(D\alpha_i^{-1}(y))|\ \|(D\alpha_i D\alpha_i^t)(\alpha_i^{-1}(y))\|_2\} \\
&\lesssim a_{max},
\end{aligned}$$

and finally the bound on the derivatives :

$$\begin{aligned}
\|DA_i(y)\|_{\mathcal{L}(\mathbb{R}^d,M_d)} &\leq a_{max}\max_{y\in\overline{Q}_+}\{\|DB_i(y)\|_{\mathcal{L}(\mathbb{R}^d,M_d)}\} + a'_{max}\max_{y\in\overline{Q}_+}\{\|D\alpha_i^{-1}(y)\|\ \|B_i(y)\|\} \\
&\lesssim a_{max} + a'_{max},
\end{aligned}$$

where we have used the notation $B_i(y) = |\det(D\alpha_i^{-1}(y))|(D\alpha_i D\alpha_i^t)(\alpha_i^{-1}(y))$. We now extend A_i to $\mathbb{R}_+^d$, for this purpose we consider two open spaces Q_i and $\tilde{Q}_i$ such that $\mathrm{supp}(v_i) \subset Q_i \subset \overline{Q}_i \subset \widetilde{Q}_i \subset \overline{\widetilde{Q}_i} \subset Q_+$ and consider $\psi \in \mathcal{C}^\infty(\mathbb{R}^d, [0,1])$ such that $\psi = 0$ on Q_i and $\psi = 1$ on $\overline{\widetilde{Q}_i}^c$. We define then $\bar{A}$ on $\mathbb{R}_+^d$ by

$$\bar{A}(x) = \begin{cases} A(x)(1-\psi(x)) + a_{min}\psi(x)I_d & \text{if } x \in Q_+ \\ a_{min}\psi(x)I_d & \text{if } x \in Q_+^c \end{cases}$$

Analogously to the case of $\bar{a}$, we have then for any $y \in \mathbb{R}_+^d$, and $z \in \mathbb{R}^d$,

$$\|\bar{A}_i(y)\| \lesssim a_{max},$$

$$\bar{A}_i(y)z.z \gtrsim a_{min}\|z\|^2,$$
$$\|D\bar{A}_i(y)\|_{\mathcal{L}(\mathbb{R}^d,M_d)} \lesssim \bar{a}'_{max} := a_{max} + a'_{max}.$$

Then we continue F_i on $\mathbb{R}^d_+$ by 0, which yields $F_i \in L^2(\mathbb{R}^d_+)$. Besides this we continue v_i on $\mathbb{R}^d_+$ by 0, which yields $v_i \in H^1_0(\mathbb{R}^d_+)$. v_i is then solution on $\mathbb{R}^d_+$ of

$$-\mathrm{div}(\bar{A}(x)\nabla v_i(x)) = -F_i(x),$$

which enables us to apply Proposition 2.9.3 to obtain that $v_i \in H^2(\mathbb{R}^d_+)$ with

$$|v_i|_{H^2(\mathbb{R}^d_+)} \lesssim \frac{a_{max}}{a_{min}}(\bar{a}'_{max}|v_i|_{H^1(\mathbb{R}^d_+)} + \|F_i\|_{L^2(\mathbb{R}^d_+)}).$$

In particular,

$$\|v_i\|_{H^2(Q_+)} \lesssim \|v_i\|_{H^1(Q_+)} + \frac{a_{max}}{a_{min}}(\bar{a}'_{max}|v_i|_{H^1(\mathbb{R}^d_+)} + \|F_i\|_{L^2(\mathbb{R}^d_+)}).$$

We define $\delta = \min_{x\in\bar{D}\cap\bar{D}_i} |\mathrm{Det}(D\alpha_i(x))|$, then using theorem 6.2.17 of [39], and recalling that $u_i(x) = v\circ\alpha_i(x)$ for any $x \in D\cap D_i$ and $\mathrm{supp}(u_i) \subset D_i$, we have $\|u_i\|_{H^2(D)} \leq \frac{C}{\sqrt{\delta}}\|\alpha_i\|_{\mathcal{C}^2(\bar{D}_i)}\|v_i\|_{H^2(Q_+)}$. We have also the elementary inequality : $\|F_i\|_{L^2(\mathbb{R}^d_+)} \leq \frac{1}{\sqrt{\delta}}\|g_i\|_{L^2(Q_+)}$. Similary, denoting $\delta' = \min_{x\in\bar{D}\cap\bar{D}_i} |\mathrm{Det}(D\alpha_i(x))|^{-1}$, we have $\|v_i\|_{H^1(Q_+)} \leq \frac{C}{\sqrt{\delta'}}\|(\alpha_i)^{-1}\|_{\mathcal{C}^2(\bar{D}_i)}\|u_i\|_{H^1(D\cap D_i)}$. Hence,

$$\begin{aligned}\|u_i\|_{H^2(D)} &\lesssim \frac{a_{max}}{a_{min}}(\bar{a}'_{max}\|u_i\|_{H^1(D)} + \|g_i\|_{L^2(D)}) + \|u_i\|_{H^1(D)}\\ &\lesssim \frac{\|f\|_{L^2(D)}}{a^3_{min}}(a_{max} + a'_{max})a_{max}.\end{aligned}$$

□

We can then conclude the proof of the theorem, recalling that $u = \sum_{i=0}^p u_i$.

Chapitre 3

Résultats numériques

Dans ce chapitre, on illustre numériquement les résultats du Chapitre 2. Tous les exemples numériques ci-dessous concernent le cas 1D d'un champ logormal a avec covariance exponentielle défini dans le paragraphe 2.7.1. On utlise les expressions explicites pour les valeurs et vecteurs propres du développement de Karhunen-Loève de $\log(a)$ vues dans la paragraphe 2.7.1.
Dans une première partie, on s'intéresse à l'erreur issue de la troncature du développement de Karhuen-Loève de $\log(a)$. On commence par s'intéresser à la décroissance des valeurs propres qui influe bien évidemment sur la vitesse de convergence de la solution u_N de l'EDP correspondant au champ tronqué a_N vers la solution u de l'EDP avec champ complet a. On retrouve numériquement l'équivalent en $1/n$ pour les valeurs propres. On s'intéresse également à l'influence des paramètres ℓ et σ sur la vitesse de décroissance des valeurs propres. On observe en particulier l'existence d'un palier de taille liée à la valeur de la longueur de corrélation ℓ avant d'atteindre cette décroissance asymptotique en $1/n$; cette taille est de l'ordre de $1/\ell$. Ensuite on s'intéresse à l'erreur provenant de la troncature sur le coefficient a, à savoir l'erreur commise en approchant a_N par a. Et enfin on s'intéresse à l'erreur faible commise sur la solution, c'est-à-dire entre u_N et u. On observe alors que l'ordre faible obtenu dans le Théorème 2.5.3 de la partie 2.5 est optimal, puisqu'on retrouve bien une erreur faible en $1/N$ dans le cas de notre exemple. On s'intéresse également à l'influence des paramètres σ et λ, retouvant bien entendu, comme pour l'étude de la décroissance des valeurs propres, une détérioration de la convergence lorsque σ augmente ou que ℓ diminue, et l'existence d'un palier avant d'atteindre une convergence asymptotique en $1/N$, la taille de ce palier étant naturellement la même que dans le cas de la décroissance des valeurs propres et donc liée à la valeur de ℓ.
Dans une seconde partie, on s'intéresse à l'erreur éléments finis, à la fois pour la solution de l'EDP avec champ complet a peu régulier ($\mathcal{C}^{1/2-\varepsilon}$ pour tout $\varepsilon > 0$) et pour la solution de l'EDP avec champ tronqué a_N qui est lui très régulier. On s'intéresse tout d'abord à l'erreur éléments finis trajectorielle, c'est-à-dire correspondant à une réalisation fixée du coefficient, puis à l'erreur éléments finis sur la loi. On retrouve numériquement les ordres attendus pour l'erreur en norme L^2, à savoir 2 pour le champ tronqué, et presque 1 pour le champ complet, montrant que les estimations obtenues dans les Propositions 4.3.10, 2.6.3 et 4.3.15 sont optimales. Il est intéressant de noter que dans le cas du champ tronqué, la convergence asymptotique en h^2 est précédée d'une convergence en h dans la zone où h est supérieur à $1/N$. On note également que dans le cas du champ tronqué, la constante multiplicative apparaissant dans l'erreur éléments finis croît naturellement avec N. Ces deux phénomènes proviennent du fait qu'à la limite, lorsque N tend vers l'infini, u_N converge vers u qui est moins régulier.
Enfin, dans une troisième partie, on s'intéresse à l'erreur de collocation étudiée dans le paragraphe 2.6. Les résultats numériques semblent indiquer une convergence plus rapide que l'estimation obtenue dans le Théorème 2.6.8, plus proche des estimations obtenues dans [1] dans le cas de variables aléatoires à support borné.

3.1 Erreur de troncature

3.1.1 Décroissance des valeurs propres

On se place en dimension 1 ; pour simplifier on se place sur $D = (0,1)$, et on considère un champ lognormal $a = e^g$ homogène de covariance exponentielle, ie $cov[g](x,y) = \sigma^2 e^{-\frac{|x-y|}{l}}$, où σ est l'écart-type et l la longueur de corrélation, comme dans le paragraphe 2.7.1. On considère le développement de Karhunen-Loève de g :

$$g(\omega, x) = \sum_{n\geq 0} \sqrt{\lambda_n} b_n(x) Y_n(w).$$

On rappelle que les Y_n sont des variables aléatoires gaussiennes centrées réduites indépendantes, et que dans notre cas, les valeurs propres λ_n et les vecteurs propres b_n de l'opérateur de covariance peuvent être calculés explicitement. Plus précisément, considérant l'équation caractéristique

$$(\ell^2 w^2 - 1)\sin(w) = 2\ell w \cos(w)$$

et notant $(w_n)_{n\geq 1}$ la suite de ces racines positives ordonnées dans l'ordre croissant, alors les valeurs propres du développement de Karhunen-Loève de g sont définies par

$$\lambda_n = \frac{2\ell\sigma^2}{\ell^2 w_n^2 + 1}$$

et les vecteurs propres par

$$b_n(x) = \alpha_n(\sin(w_n x) + \ell w_n \cos(w_n x)),$$

où $\alpha_n = \frac{1}{\sqrt{(\ell^2 w_n^2+1)/2+\ell}}$. On a naturellement $\sum_{n\geq 0} \lambda_n = \sigma^2$.

La suite des racines (w_n) vérifie $w_n \underset{n\to+\infty}{\sim} n\pi$, ce qui donne les équivalents suivants :

$$\lambda_n \underset{n\to+\infty}{\sim} \frac{2\sigma^2}{\ell\pi^2 n^2}$$

$$b_n(x) \underset{n\to+\infty}{\sim} \sqrt{2}\cos(w_n x).$$

On s'intéresse à l'influence des paramètres σ et ℓ. Pour commencer on remarque que $\sigma^2 \mapsto (\lambda_n)_{n\in\mathbb{N}}$ est linéaire et que les vecteurs propres (b_n) ne dépendent pas de σ. La convergence des valeurs propres se ralentit donc quand σ augmente, et cette dégradation est explicitement quantifiée par le fait que les valeurs propres sont proportionnelles à σ^2. La dépendance en ℓ est moins évidente. Tout d'abord, il est clair que $1/\ell$ apparaît comme facteur multiplicatif dans l'équivalent de λ_n, ce qu'on peut observer sur la Figure 3.1. La vitesse de convergence se dégrade donc quand ℓ diminue.

Par ailleurs, on observe numériquement (voir [12] par exemple) sur la Figure 3.1, un palier dans la décroissance des valeurs propres λ_n. En effet, on observe numériquement que la décroissance asymptotique $\lambda_n \underset{n\to+\infty}{\sim} \frac{2\sigma^2}{\ell\pi^2 n^2}$ n'est atteinte qu'après un palier de taille un peu inférieure à $1/l$, et les valeurs propres sont quasiment constantes au début du palier. En effet on peut voir sur la Figure 3.2 que dans le cas $\ell = 0,1$ le palier comporte environ 8 points, et dans le cas $\ell = 0,01$ le palier comporte environ 70 points.

3.1.2 Erreur de troncature sur le coefficient

On considère alors le développement de Karhunen-Loève tronqué :

$$g_N(\omega, x) = \sum_{n=0}^{N} \sqrt{\lambda_n} b_n(x) Y_n(w),$$

et l'approximation du champ a correspondante $a_N = e^{g_N}$. On s'intéresse ici à l'erreur forte commise en approchant a_N par a, en calculant $\|a_N - a\|_{L^2(\Omega\times(0,1))}$. Il découle directement de la Proposition 2.3.11 que cette erreur peut être majorée par une constante multipliée par $N^{-1/2+\varepsilon}$, pour tout $\varepsilon > 0$. On retrouve numériquement une convergence en $1/\sqrt{N}$ dans la Figure 3.3, montrant que l'estimation est optimale.

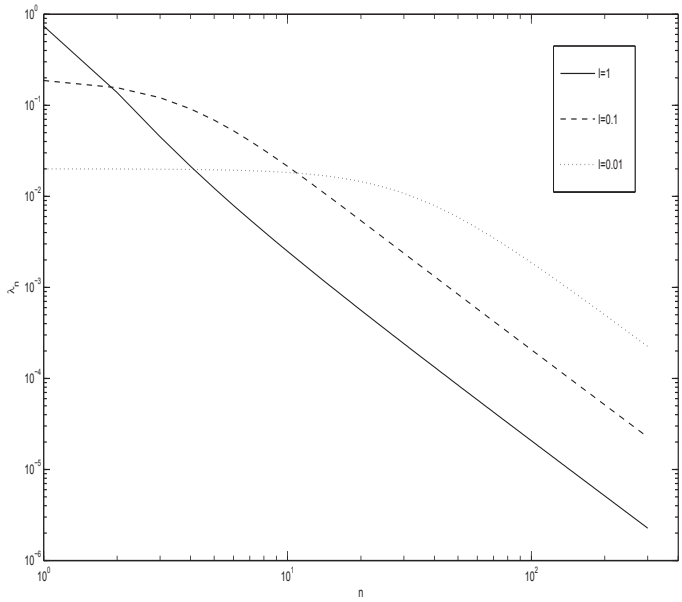


FIG. 3.1 – λ_n en fonction de n, en échelle logarithmique, pour $\sigma = 1$ et différentes valeurs de l .

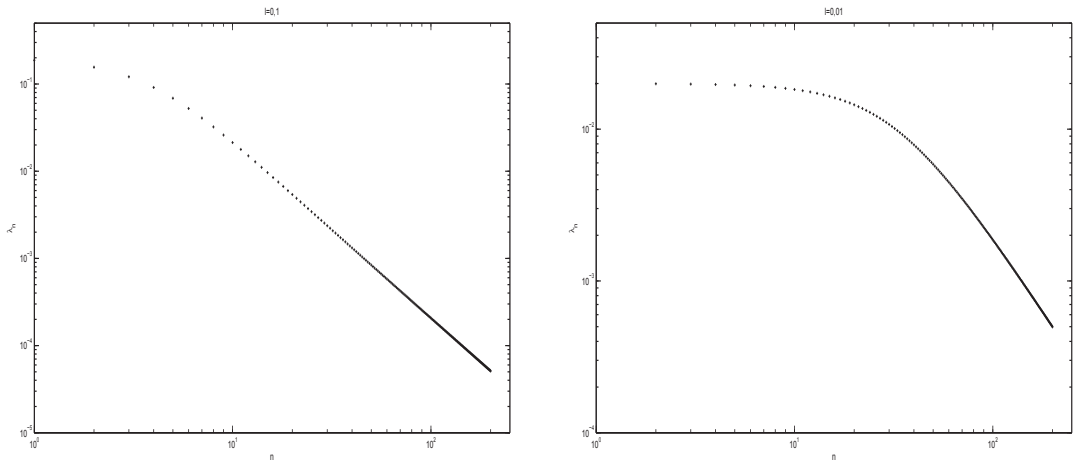


FIG. 3.2 – λ_n en fonction de n en échelle logarithmique, pour $\sigma = 1$ et $l = 0,1$ (à gauche) et $l = 0,01$ (à droite) .

3.1.3 Erreur faible de troncature sur la solution

On note u_N et u les solutions des EDP correspondantes, où on a considéré le cas de l'équation avec conditions de Dirichlet homogènes et second membre égal à 1, c'est-à-dire que u est définie comme la solution de :

$$\begin{cases} -\mathrm{div}(a(\omega,x)\nabla u(\omega,x)) = 1 \text{ sur } (0,1), \\ u(0) = 0,\ u(1) = 0, \end{cases} \tag{3.1}$$

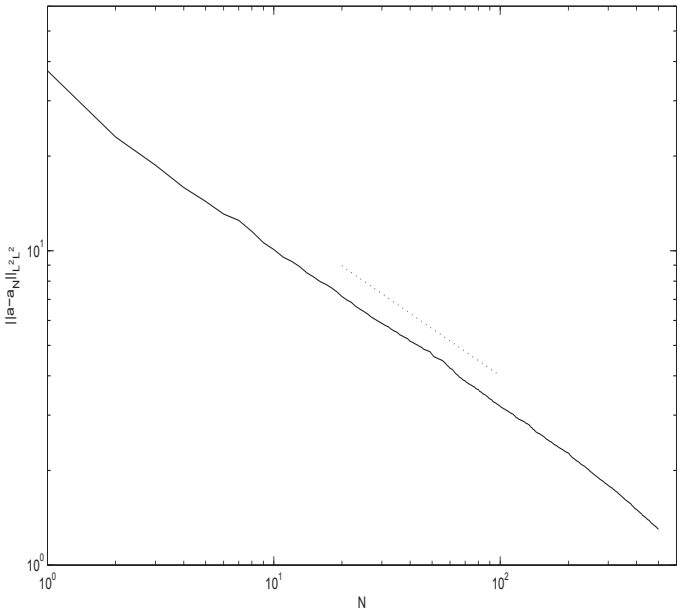


FIG. 3.3 – $\|a_N - a\|_{L^2(\Omega\times(0,1))}$ en fonction de N, en échelle logarithmique, pour $\sigma = 1$ et $l = 1$. Les pointillés indiquent une pente de $-1/2$.

et u_N et définie comme la solution de :

$$\begin{cases} -\mathrm{div}(a_N(\omega,x)\nabla u_N(\omega,x)) = 1 \text{ sur } (0,1), \\ u_N(0) = 0,\ u_N(1) = 0. \end{cases} \tag{3.2}$$

On s'intéresse ici à l'erreur faible (c'est-à-dire en loi) commise en approchant la solution u de (3.1) par u_N la solution de (3.2). Pour commencer on va donc comparer $x \mapsto \mathbb{E}[\varphi(u_N)](x)$ à $x \mapsto \mathbb{E}[\varphi(u)](x)$, pour des fonctions φ régulières, $\varphi : \mathbb{R} \to \mathbb{R}$. On utilise une méthode de collocation stochastique (avec grilles pleines) pour calculer les quantités $x \mapsto \mathbb{E}[\varphi(u_N)]$ et une méthode de Monte-Carlo pour calculer les quantités $x \mapsto \mathbb{E}[\varphi(u)](x)$. La taille de la grille d'éléments finis, le nombre de points de collocation et le nombre de réalisations dans la méthode de Monte-Carlo soit choisis dans chaque cas de telle sorte que les erreurs correspondantes soient négligeables par rapport à l'erreur de troncature. Les Figures 3.4 et 3.5 montrent des exemples de calculs de $x \mapsto \mathbb{E}[u_N]$ et $x \mapsto \mathbb{E}[u](x)$. On peut déjà remarquer que la convergence est très rapide dans le cas $\ell = 1$, dans ce cas il suffit de prendre quelques termes dans le développement de Karhunen-Loève pour avoir une bonne approximation de l'espérance. La convergence est moins rapide dans le cas $\ell = 0,1$.

On va maintenant s'intéresser à la vitesse de convergence de $x \mapsto \mathbb{E}[u_N]$ vers $x \mapsto \mathbb{E}[u](x)$. On a montré dans le paragraphe 2.5, que si $\varphi \in \mathcal{C}^4(\mathbb{R},\mathbb{R})$ est à dérivées bornées, alors $\|\mathbb{E}[\varphi(u_N) - \varphi(u)]\|_{L^p(D)} \leq C_{2.7.2}(\varphi,p)\frac{1}{N}$, d'après la Proposition 2.7.2 déduite du Théorème 2.5.3. On évalue maintenant numériquement la vitesse à laquelle $\mathbb{E}[u_N]$ converge vers $\mathbb{E}[u]$ dans notre cas, en traçant le logarithme de $\|\mathbb{E}[u-u_N]\|_{L^2}$ en fonction du logarithme de N.

On retouve bien dans la Figure 3.6 une convergence en $1/N$ dans les deux cas, $l = 1$ et $l = 0,1$, ce qui montre que notre estimation d'erreur faible de troncature est optimale. On peut voir, toujours sur la Figure 3.6, que la constante multiplicative C est plus importante dans le cas $l = 0,1$ et donc l'erreur plus importante, comme on pouvait s'y attendre, et dans le cas $l = 0.1$ on retrouve également un plateau avant d'atteindre la convergence asymptotique en $1/N$, qui correspond au plateau dans la décroissance des valeurs propres de la Figure 3.2.

On s'intéresse maintenant au cas où $\varphi : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$, c'est-à-dire au cas où l'on souhaite calculer

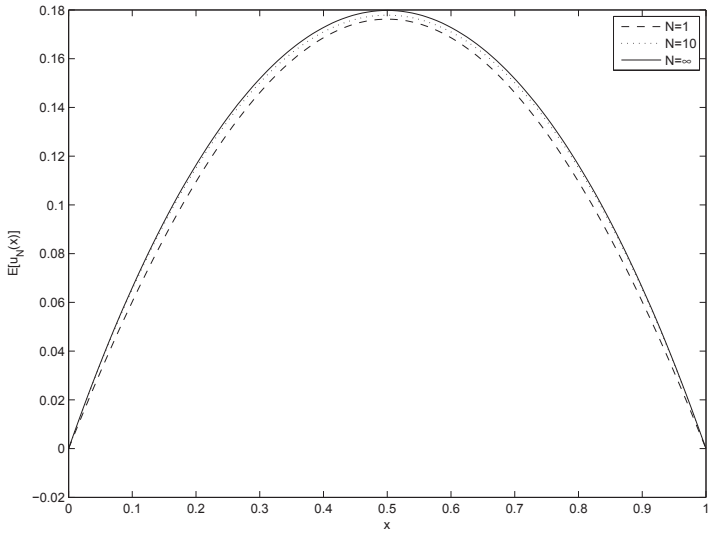


FIG. 3.4 – $\mathbb{E}[u_N]$ en fonction de x, pour $\sigma = 1, l = 1$ et différentes valeurs de N.

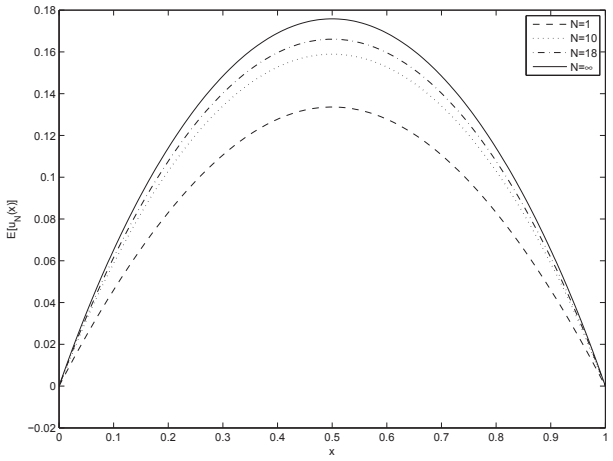


FIG. 3.5 – $\mathbb{E}[u_N]$ en fonction de x, pour $\sigma = 1, l = 0.1$ et différentes valeurs de N.

$\mathbb{E}[u^2](x)$. On observe sur la Figure 3.7 qu'on obtient une erreur faible en $1/N$ pour le moment d'ordre deux comme pour l'espérance ; la convergence est néanmoins plus lente pour le moment d'ordre deux que pour l'espérance.

Pour conclure, on s'intéresse à une forme d'erreur faible un peu différente. On considère $\psi : H_0^1(0,1) \to \mathbb{R}$, $u \mapsto \|u\|_{L^2(0,1)}$, et on s'intéresse à l'erreur commise en approchant $\mathbb{E}[\|u\|_{L^2(0,1)}]$ par $\mathbb{E}[\|u_N\|_{L^2(0,1)}]$. Comme précédemment, $\mathbb{E}[\|u\|_{L^2(0,1)}]$ est calculée à l'aide d'une méthode de Monte-Carlo et $\mathbb{E}[\|u_N\|_{L^2(0,1)}]$ à l'aide d'une méthode de collocation stochastique. On peut voir sur les Figures 3.8 et 3.9 qu'on retrouve dans chaque cas une erreur faible en $1/N$ comme on pouvait s'y attendre au vu du Théorème 2.5.3 et de la Proposition

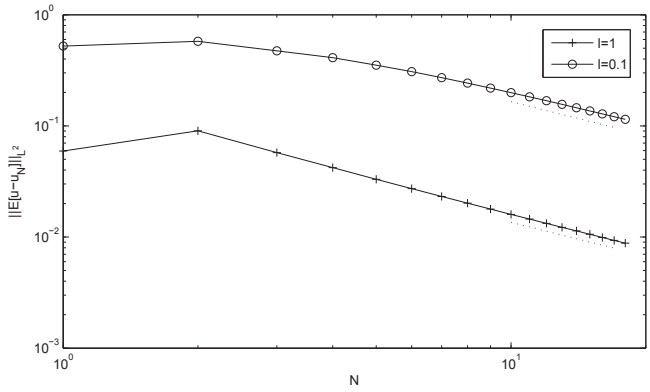


FIG. 3.6 – $\|\mathbb{E}[u-u_N]\|_{L^2}$ en fonction de N, en échelle logarithmique, pour $\sigma = 1$ dans les cas $l = 1$ et $l = 0,1$. Les pointillés indiquent une pente de -1.

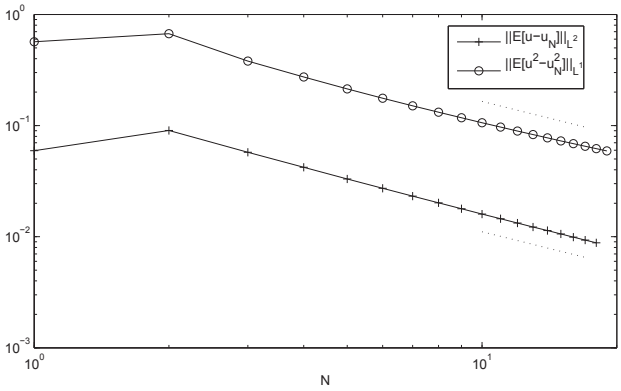


FIG. 3.7 – $\|\mathbb{E}[u-u_N]\|_{L^2}$ et $\|\mathbb{E}[u^2-u_N^2]\|_{L^1}$ en fonction de N, en échelle logarithmique, pour $\sigma = 1$ et $l = 1$. Les pointillés indiquent une pente de -1.

2.7.2 qui en découle, même si le type d'erreur faible qu'on calcule ici ne rentre pas dans le cadre des hypothèses de ces résultats théoriques. On s'intéresse par ailleurs encore une fois à l'influence des valeurs de l et de σ. Sur la Figure 3.8, on peut voir comme on l'a déjà vu que la convergence se détériore quand ℓ diminue et que la convergence asymptotique en $1/N$ n'est atteinte qu'après un palier de taille un peu inférieure à $1/\ell$, à savoir ici pour $\ell = 0,1$ à partir de $N = 8$ environ. Sur la figure 3.9, on peut voir que la convergence se détériore rapidement quand σ augmente, en effet pour $\sigma = 2$ l'erreur est environ multipliée par un facteur 30 par rapport au cas $\sigma = 1$.

3.2 Erreur éléments finis

Dans cette partie, on s'intéresse à l'erreur éléments finis, à la fois dans le cas du champ complet a et dans le cas du champ tronqué a_N. Comme on l'a vu dans la Proposition 2.2.1, les réalisations de a sont $1/2-\varepsilon$-Hölderiennes pour tout $\varepsilon > 0$; on en déduira dans le Théorème 4.3.9 et le Corollaire 4.3.10 des estimations de l'erreur éléments finis en norme L^2, trajectorielle (c'est-à-dire à ω fixé) et forte (plus précisément en

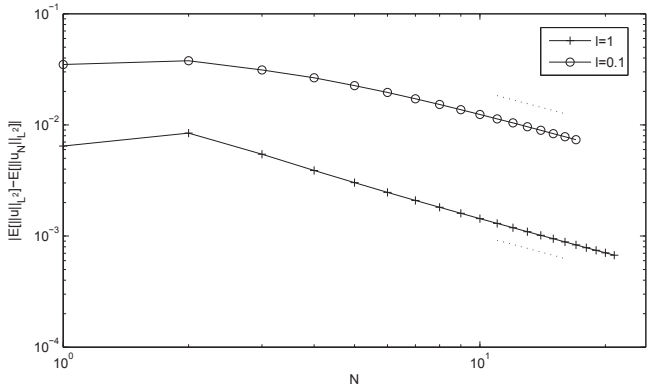


FIG. 3.8 – $|\mathbb{E}[\|u\|_{L^2}] - \mathbb{E}[\|u_N\|_{L^2}]|$ en fonction de N, en échelle logarithmique, pour $\sigma = 1$ dans les cas $l = 1$ et $l = 0, 1$. Les pointillés indiquent une pente de -1.

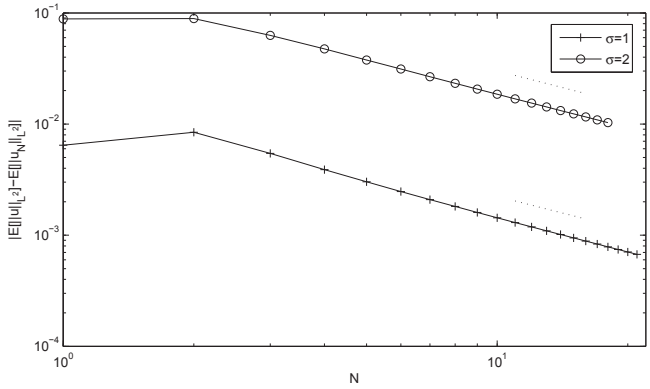


FIG. 3.9 – $|\mathbb{E}[\|u\|_{L^2}] - \mathbb{E}[\|u_N\|_{L^2}]|$ en fonction de N, en échelle logarithmique, pour $l = 1$ dans les cas $\sigma = 1$ et $\sigma = 2$. Les pointillés indiquent une pente de -1.

norme $L^p(\Omega, L^2(D)))$ avec un ordre $1 - \varepsilon$ pour tout $\varepsilon > 0$. Les réalisations de a_N sont quant à elles très régulières ($\mathcal{C}^\infty$). On a donc une estimation d'erreur éléments finis en norme L^2 théorique en $c_N h^2$, comme on peut le déduire de la Proposition 2.6.3. On verra dans le Théorème 4.3.15 qu'on peut également avoir une estimation d'erreur éléments finis en norme L^2 indépendante de N de la forme $ch^{1-\varepsilon}$ pour tout $\varepsilon > 0$, c'est-à-dire avec l'ordre correspondant à la régularité de la limite u de u_N. Toutes ces estimations d'erreurs éléments finis sont des estimations d'erreur fortes (plus précisément en norme $L^p(\Omega, L^2(D)))$ découlant d'estimations trajectorielles (c'est-à-dire à ω fixé). On peut donc bien évidemment majorer les erreurs d'éléments finis sur la loi par les erreurs fortes. Dans ce paragraphe, tous les résultats numériques concernent le cas $\sigma = 1$ et $\ell = 1$.

3.2.1 Erreur éléments finis trajectorielle

On commence par s'intéresser à l'erreur éléments finis pour une réalisation du coefficient. Tout d'abord on considère la cas du coefficient complet a et on s'intéresse à l'erreur commise en approchant $u(\omega, \cdot)$ par $u^h(\omega, \cdot)$. D'après ce qui précède, on s'attend à un ordre presque 1 en norme L^2, ce qu'on retrouve numériquement sur

la Figure 3.10, montrant l'optimalité de l'estimation d'erreur du Théorème 4.3.15 .

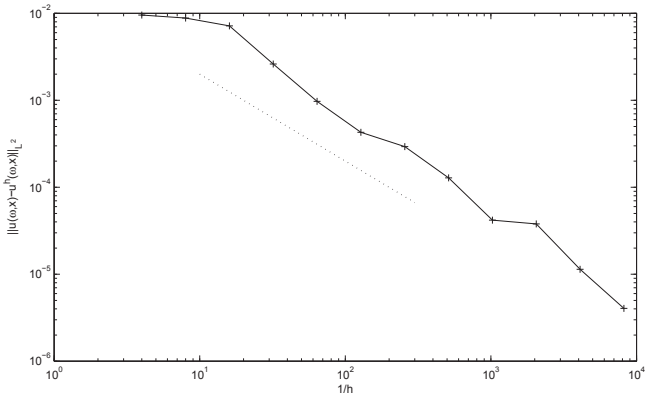


FIG. 3.10 – $\|u(\omega,x) - u^h(\omega,x)\|_{L^2(0,1)}$ en fonction de h, en échelle logarithmique. Les pointillés indiquent une pente de -1.

On considère ensuite le cas du coefficient tronqué a_N, on s'intéresse alors à l'erreur commise en approchant $u_N(\omega,\cdot)$ par $u_N^h(\omega,\cdot)$. D'après ce qui précède, on s'attend à un ordre 2 d'après la Proposition 2.6.3, ce qu'on retrouve numériquement sur la Figure 3.11. Néanmoins, il est intéressant de noter que la convergence asymptotique en h^2 n'est obtenue que pour h suffisament petit, environ à partir de $1/N$. En effet on peut voir sur la courbe de gauche de la Figure 3.11 qui correspond au cas $N = 500$ que la convergence asymptotique d'ordre 2 commence environ pour $1/h$ supérieur à 500, et on peut voir sur la courbe de droite de la Figure 3.11 qui correspond au cas $N = 2000$ que la convergence asymptotique d'ordre 2 commence environ pour $1/h$ supérieur à 2000. On remarque dans les deux courbes de la Figure 3.11 qu'avant d'atteindre l'ordre asymptotique, on a un ordre 1 environ. Ceci est cohérent avec le fait qu'à la limite, quand N tend vers l'infini, le coefficient n'est pas aussi régulier et l'ordre de convergence est plus faible (presque égal à 1, voir ci-dessus). Le fait que cette convergence asymptotique commence environ à partir de $1/N$ peut s'expliquer par le fait que a_N "ressemble" à une somme de termes en $\cos(n\pi x)$ pour $1 \leq n \leq N$ et donc une grille d'éléments finis de pas supérieur à $1/N$ voit peu la différence entre u_N et u, alors qu'une grille de pas inférieur à $1/N$ voit u_N comme très régulière.

3.2.2 Erreur éléments finis sur la loi

On s'intéresse maintenant à l'erreur éléments finis sur la loi pour le champ tronqué, c'est-à-dire à l'erreur commise en approchant $\mathbb{E}[\varphi(u_N)]$ par $\mathbb{E}[\varphi(u_N^h)]$ Cette quantité est approchée à l'aide d'une méthode de collocation stochastique (avec grille pleine). Sous des hypothèses sur φ (hypothèses de régularité et caractère borné des dérivées) on peut déduire facilement des estimations d'erreurs éléments finis faibles à partir des estimations d'erreurs fortes évoquées ci-dessus. On s'intéresse cas du calcul de l'espérance, c'est-à-dire à l'approximation de $\mathbb{E}[u_N]$ par $\mathbb{E}[u_N^h]$, dans le cas $\sigma = 1$, $\ell = 1$.

On obtient donc numériquement une erreur faible de la forme $c_N h^2$ d'après la Figure 3.12 comme attendu au vu de l'estimation d'erreur forte de la Proposition 2.6.3. Comme ci-dessus, on peut voir sur la Figure 3.12 que cet ordre asymptotique n'est atteint que pour $1/h$ supérieur à environ N. On peut également voir sur la Figure 3.12 et plus précisément dans le Tableau 3.13 que la constante c_N augmente avec N, ce qui semble naturel, puisque comme on l'a déjà vu, à la limite N tendant vers l'infini, le coefficient n'est pas aussi régulier et l'ordre de convergence est plus faible .

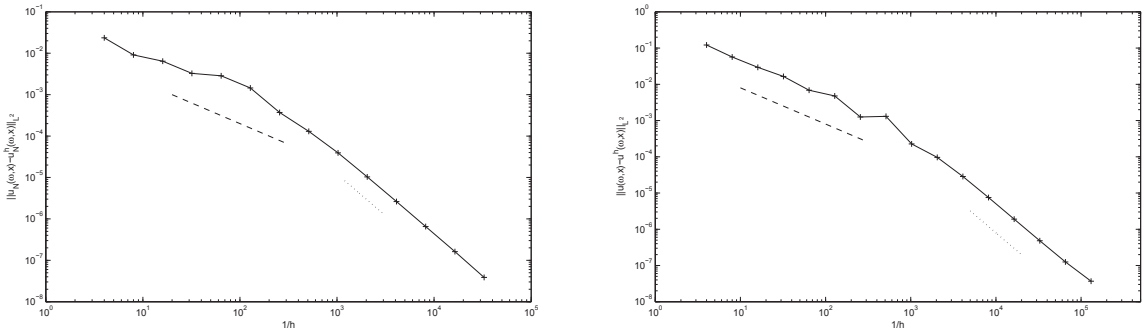


FIG. 3.11 – $\|u_N(\omega,x)-u_N^h(\omega,x)\|_{L^2(0,1)}$ en fonction de h, en échelle logarithmique, pour $N=500$ (à gauche) et $N=2000$ (à droite). Les pointillés indiquent une pente de -2, les pointillés larges indiquent une pente de -1.

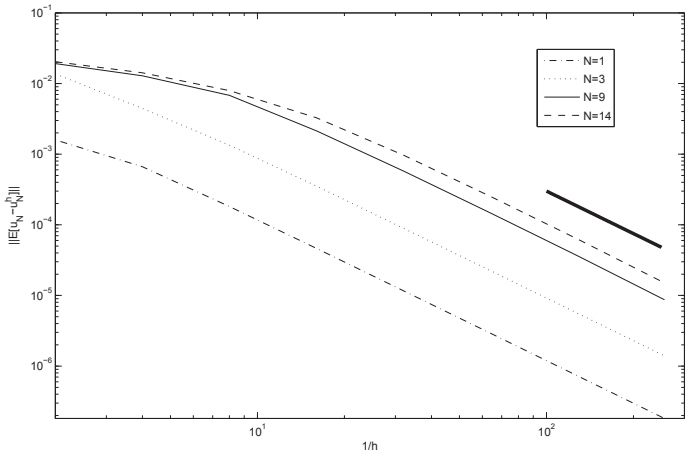


FIG. 3.12 – $\|\mathbb{E}[u_N^h]-\mathbb{E}[u_N]\|_{L^2}$ en fonction de h, pour $N=1,3,9,14$. Le trait épais indique une pente de -2.

3.3 Erreur collocation

On s'intéresse ici à l'erreur commise en approchant $\mathbb{E}[(u_N)^2]$ par la méthode de collocation stochastique, pour N fixé. Plus précisément, on prend $N=10$, $l=1$, $\sigma=1$, on fixe le pas de maillage $h=1/50$ et on estime l'erreur commise en approchant $\mathbb{E}[(u_N^h)^2]$ par $\mathbb{E}[(u_N^{h,p})^2]$ en fonction de $p=(p_1,...p_N)$, où p_i est le nombre de points de collocation dans la i-ième direction. On fixe tous les p_i sauf un, que l'on fait varier de 1 à 10.

On a obtenu dans la Proposition 2.6.8, une estimation de l'erreur de collocation de la forme

$$\sum_{i=1}^{N} C_i\sqrt{p_i}e^{-k_i\sqrt{p_i}}.$$

Les résultats numériques de la Figure 3.14 semblent plutôt indiquer que l'erreur $\|\mathbb{E}[(u_N^{h,p})^2]-\mathbb{E}[(u_N^h)^2]\|_{L^1}$ est environ de la forme $\sum_{i=1}^{N} C_i e^{-k_i p_i}$, pour les valeurs de i auxquelles on s'intéresse, à savoir seulement les

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
c_N	0,01	0,02	0,08	0,17	0,28	0,37	0,46	0,55	0,64	0,72	0,80	0,88	0,95	1,02

FIG. 3.13 – Valeurs de c_N dans le cas $\sigma = 1$, $\ell = 1$ pour N variant de 1 à 14.

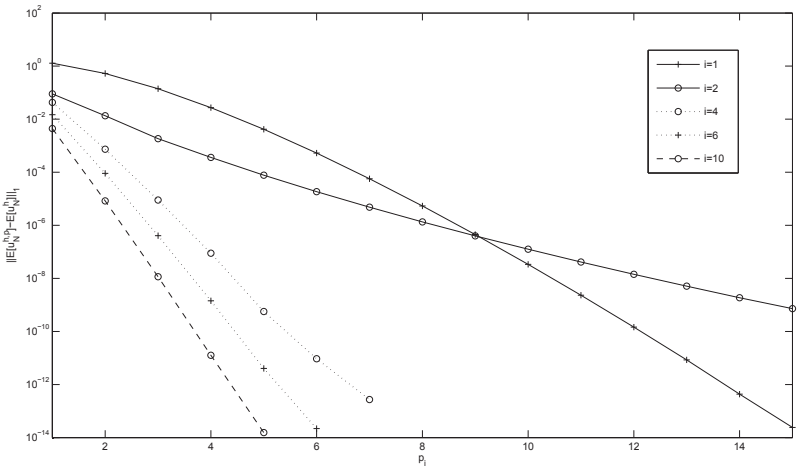


FIG. 3.14 – $\|\mathbb{E}[(u_N^{h,p})^2] - \mathbb{E}[(u_N^h)^2]\|_{L^1}$ en fonction de p_i, en échelle semi-logarithmique, pour $i = 1, 2, 4, 6, 10$.

premières directions. Les valeurs des constantes k_i et C_i sont alors données dans le Tableau 3.15.

Dans le cas des deux premières directions ($i = 1, 2$), l'erreur n'est sous cette forme que pour p_i supérieur à 5. Le fait qu'on semble obtenir pour les valeurs de i considérées (c'est-à-dire N assez faible) une convergence plus rapide que celle attendue d'après le résultat de la Proposition 2.6.8 (qui étend le résultat obtenu dans [1]) pourrait s'interpréter grâce au résultat d'estimation d'erreur de collocation obtenu dans [1] dans le cas de variables aléatoires Y_n à support borné, sous l'Hypothèse 1.1.1. En effet dans ce cas, on a une estimation d'erreur de la forme $\sum_{i=1}^{N} C_i e^{-k_i p_i}$, et on peut penser que les variables aléatoires gaussiennes Y_n "ressemblent" numériquement à des variables aléatoires à support borné.

i	1	2	3	4	5	6	7	8	9	10
c_i	$1,4\ 10^4$	$4\ 10^{-3}$	1,1	2,2	5,8	4,5	2,4	4,5	4,7	4,0
k_i	2,6	1,0	3,1	4,2	5,1	5,5	5,6	6,2	6,5	6,6

FIG. 3.15 – Valeurs des k_i et c_i dans le cas $\ell = 1$ et $\sigma = 1$ pour $N = 10$

Deuxième partie

Analyse numérique de l'erreur éléments finis pour des EDP elliptiques à coefficients aléatoires. Application aux méthodes de Monte-Carlo multi-niveaux

Chapitre 4

Finite Element Error Analysis of Elliptic PDEs with Random Coefficients and its Application to Multilevel Monte Carlo Methods

Abstract : We consider a finite element approximation of elliptic partial differential equations with random coefficients. Such equations arise, for example, in uncertainty quantification in subsurface flow modelling. Models for random coefficients frequently used in these applications, such as log-normal random fields with exponential covariance, have only very limited spatial regularity, and lead to variational problems that lack uniform coercivity and boundedness with respect to the random parameter. In our analysis we overcome these challenges by a careful treatment of the model problem almost surely in the random parameter, which then enables us to prove uniform bounds on the finite element error in standard Bochner spaces. These new bounds can then be used to perform a rigorous analysis of the multilevel Monte Carlo method for these elliptic problems that lack full regularity and uniform coercivity and boundedness. To conclude, we give some numerical results that confirm the new bounds.

Keywords : PDEs with stochastic data, non-uniformly elliptic, non-uniformly bounded, lack of full regularity, log-normal coefficients, truncated Karhunen-Loève expansion.

Résumé : On s'intéresse à l'approximation par éléments finis d'équations aux dérivées partielles elliptiques à coefficients aléatoires. De telles équations sont utilisées par exemple dans la modélisation de la quantification des incertitudes pour les écoulements souterrains. Les modèles fréquemment utilisés dans ce type d'applications, tels que des champs lognormaux à covariance exponentielle, ont une régularité spatiale faible et conduisent à des formaulations variationnelles qui ne sont pas uniformément coercives ni bornées par rapport au paramètre aléatoire. Dans notre analyse, nous surmontons ces difficultés grâce à une analyse précise du problème modèle réalisée presque sûrement par rapport au paramètre aléatoire, ce qui nous permet d'obtenir des majorations pour l'erreur éléments finis dans des espaces de Bochner standards. Ces nouvelles majorations sont utilisées pour obtenir une analyse rigoureuse de la méthode de Monte-Carlo multi-niveaux pour ces EDP elliptiques avec coefficients non uniformément bornés et coercifs. Pour conclure, on donne des résultats du numériques qui confirment ces nouvelles estimations d'erreur.

Mots clés : EDP avec coefficients aléatoires, non-uniformément elliptique, non-uniformément borné, faible régularité, coefficients lognormaux, développement de Karhunen-Loève tronqué.

4.1 Introduction

Partial differential equations (PDEs) with random coefficients are commonly used as models for physical processes in which the input data are subject to uncertainty. It is often of great importance to quantify the uncertainty in the outputs of the model, based on the information that is available on the uncertainty of the input data.

In this paper, we consider elliptic PDEs with random coefficients, as they arise for example in subsurface flow modelling (see e.g. [17], [16]). The classical equations governing a steady state, single phase subsurface flow consist of Darcy's law coupled with an imcompressibility condition. Taking into account the uncertainties in the source terms f and the permeability a of the medium, this leads to a linear, second-order elliptic PDE with random coefficient $a(\omega, x)$ and random right hand side $f(\omega, x)$, subject to appropriate boundary conditions.

Solving equations like this numerically can be challenging for several reasons. Models typically used for the coefficient $a(\omega, x)$ in applications can vary on a fine scale and have relatively large variances, meaning that in many cases we only have very limited spatial regularity. In the context of subsurface flow modelling, for example, a model frequently used for the permeability $a(\omega, x)$ is a homogeneous log-normal random field. That is $a(\omega, x) = \exp[g(\omega, x)]$, where g is a Gaussian random field. We show in §4.2.3 that for common choices of mean and covariance functions, in particular an exponential covariance function for g, trajectories of this type of random field are only Hölder continuous with exponent less than $1/2$. Another difficulty sometimes associated with PDEs with random coefficients, is that the coefficients cannot be bounded uniformly in the random parameter ω. In the worst case, this leads to elliptic differential operators which are not uniformly coercive or uniformly bounded. This is for example true for log-normal random fields $a(\omega, x)$. Due to the nature of Gaussian random variables, $a(\omega, x)$ can in this case not be bounded from above or away from zero uniformly in ω. It is, however, possible to bound $a(\omega, x)$ for each fixed ω.

In this paper, we consider a finite element approximation (in space) of elliptic PDEs with random coefficients as described above, with particular focus on the cases where the coefficient $a(\omega, x)$ can not be bounded from above or away from zero uniformly in ω, and trajectories of $a(\omega, x)$ are only Hölder continuous. Indeed, if one assumes that the random coefficient $a(\omega, x)$ is sufficiently regular and can be bounded from above and away from zero uniformly in ω, the resulting variational problem is uniformly coercive and bounded, and the well-posedness of the problem and the subsequent error analysis are classical, see eg. [2, 24, 1, 4]. We here derive bounds on the finite element error in the solution in both $L^p(H_0^1(D))$ and $L^p(L^2(D))$ norms. Our error estimate crucially makes use of the observation that for each *fixed* ω, we have a uniformly coercive and bounded problem (in x). The derivation of the error estimate is then based on an elliptic regularity result for coefficients supposed to be only Hölder continuous, making the dependence of all constants on $a(\omega, x)$ explicit. We emphasise that we work in standard Bochner spaces with the usual test and trial spaces as in the deterministic setting. As such, our work builds on and complements [26, 10, 36, 66] which all are concerned with the well-posedness and numerical approximation of elliptic PDEs with infinite dimensional stochastic coefficients that are not uniformly bounded from above and below (e.g. log-normal coefficients).

Finally, applying the new finite element error bounds, we quantify the error committed in the multilevel Monte Carlo (MLMC) method. The MLMC analysis is motivated by [12], where the authors recently demonstrated numerically the effectiveness of MLMC estimators for computing moments of various quantities of interest related to the solution $u(\omega, x)$ of an elliptic PDE with log-normal random coefficients. The MLMC method is a very efficient variance reduction technique for classical Monte Carlo, especially in the context of differential equations with stochastic data and stochastic differential equations. It was first introduced by Heinrich [40] for the computation of high-dimensional, parameter-dependent integrals, and has been analysed extensively by Giles [35, 34] in the context of stochastic differential equations in mathematical finance. Under the assumptions of uniform coercivity and boundedness, as well as full regularity, convergence results for the MLMC estimator for elliptic PDEs with random coefficients have recently also been proved in [4]. Here we consider the problem without these assumptions which is of particular interest in subsurface flow applications, where log-normal coefficients are most commonly used.

The outline of the rest of this paper is as follows. In §2, we present our model problem and assumptions, and recall from [10], the main results on the error resulting from truncating the Karhunen-Loéve expansion in the context of log-normal coefficients. §3 is devoted to the establishment of the finite element error bound.

We first prove a regularity result for elliptic PDEs with Hölder continuous coefficients, making explicit the dependence of the bound on ω. The full proof is given in the appendix. From this regularity result we then deduce the finite element error bound in the H^1-seminorm, before also treating the case of the L^2-norm and the contribution of the quadrature error. We also improve the bound given in [10] for the finite element error in the case of truncated KL-expansions of log-normal random fields, ensuring uniformity in K (the number of terms of the truncated expansion). In §4, we use the finite element error analysis from §3 to furnish a complete error analysis of the multilevel Monte Carlo method. To begin with, we present briefly the multilevel Monte-Carlo method for elliptic PDEs with random coefficients proposed in [12], before providing a rigorous bound for the ε-cost. Finally, in §5 we present some numerical experiments, illustrating the sharpness of some of the results.

4.2 Preliminaries

4.2.1 Notation

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a bounded Lipschitz domain $D \subset \mathbb{R}^d$, we introduce the following notation. For any $k \in \mathbb{N}$, we define on the Sobolev space $H^k(D)$ the following semi-norm and norm :

$$|v|_{H^k(D)} = \left(\int_D \sum_{|\alpha|=k} |D^\alpha v|^2 \, \mathrm{d}x \right)^{1/2} \quad \text{and} \quad \|v\|_{H^k(D)} = \left(\int_D \sum_{|\alpha|\leq k} |D^\alpha v|^2 \, \mathrm{d}x \right)^{1/2}.$$

We recall that, since D is bounded, the semi-norm $|\cdot|_{H^k(D)}$ defines a norm equivalent to the norm $\|\cdot\|_{H^k(D)}$ on the subspace $H_0^k(D)$ of $H^k(D)$. For any real $r \geq 0$, with $r \notin \mathbb{N}$, set $r = k + s$ with $k \in \mathbb{N}$ and $0 < s < 1$, and denote by $|\cdot|_{H^r(D)}$ and $\|\cdot\|_{H^r(D)}$ the Sobolev–Slobodetskii semi-norm and norm, respectively, defined for $v \in H^k(D)$ by

$$|v|_{H^r(D)} = \left(\iint_{D\times D} \sum_{|\alpha|=k} \frac{[D^\alpha v(x) - D^\alpha v(y)]^2}{|x-y|^{d+2s}} \, \mathrm{d}x \, \mathrm{d}y \right)^{1/2} \text{and} \quad \|v\|_{H^r(D)} = \left(\|v\|^2_{H^k(D)} + |v|^2_{H^r(D)} \right)^{1/2}.$$

The Sobolev space $H^r(D)$ is then defined as the space of functions v in $H^k(D)$ such that the integral $|v|^2_{H^r(D)}$ is finite. For $0 < s \leq 1$, the space $H^{-s}(D)$ denotes the dual space to $H_0^s(D)$ with the dual norm.

In addition to the above Sobolev spaces, we also make use of the Hölder spaces $\mathcal{C}^t(\overline{D})$, with $0 < t < 1$, on which we define the following semi-norm and norm

$$|v|_{\mathcal{C}^t(\overline{D})} = \sup_{x,y\in\overline{D}:\, x\neq y} \frac{|v(x) - v(y)|}{|x-y|^t} \quad \text{and} \quad \|v\|_{\mathcal{C}^t(\overline{D})} = \sup_{x\in\overline{D}} |v(x)| + |v|_{\mathcal{C}^t(\overline{D})}.$$

The spaces $\mathcal{C}^0(\overline{D})$ and $\mathcal{C}^1(\overline{D})$ are as usual the spaces of continuous and continuously differentiable functions with the standard norms.

Finally, we will also require spaces of Bochner integrable functions. To this end, let $\mathcal{B}$ be a Banach space with norm $\|\cdot\|_{\mathcal{B}}$, and $v : \Omega \to \mathcal{B}$ be strongly measurable. With the norm $\|\cdot\|_{L^p(\Omega,\mathcal{B})}$ defined by

$$\|v\|_{L^p(\Omega,\mathcal{B})} = \begin{cases} \left(\int_\Omega \|v\|^p_{\mathcal{B}} \, \mathrm{d}\mathbb{P}\right)^{1/p}, & \text{for} \quad p < \infty, \\ \operatorname{esssup}_{\omega\in\Omega} \|v\|_B, & \text{for} \quad p = \infty, \end{cases}$$

the space $L^p(\Omega, \mathcal{B})$ is defined as the space of all strongly measurable functions on which this norm is finite. In particular, we denote by $L^p(\Omega, H_0^k(D))$ the Bochner space where the norm on $H_0^k(D)$ is chosen to be the seminorm $|\cdot|_{H^k(D)}$. For simplicity we write $L^p(\Omega)$ for $L^p(\Omega, \mathbb{R})$.

The key task in this paper is keeping track of how the constants in the bounds and estimates depend on the coefficient $a(\omega, x)$ and on the mesh size h. Hence, we will almost always be stating constants explicitly. To simplify this task we shall denote constants appearing in Theorem/Proposition/Corollary x.x by $C_{x.x}$. Constants that do not depend on $a(\omega, x)$ or h will not be explicitly stated. Instead, we will write $b \lesssim c$ for two positive quantities b and c, if b/c is uniformly bounded by a constant independent of $a(\omega, x)$ and of h.

4.2.2 Problem Setting

We consider the following linear elliptic partial differential equation (PDE) with random coefficients :

$$\begin{aligned} -\nabla \cdot (a(\omega,x)\nabla u(\omega,x)) &= f(\omega,x), && \text{in} \quad D, \\ u(\omega,x) &= 0, && \text{on} \quad \partial D, \end{aligned} \tag{4.1}$$

for almost all $\omega \in \Omega$. The differential operators $\nabla\cdot$ and ∇ are with respect to $x \in D$. Let us formally define, for all $\omega \in \Omega$,

$$a_{\min}(\omega) := \min_{x\in\overline{D}} a(\omega,x) \qquad \text{and} \qquad a_{\max}(\omega) := \max_{x\in\overline{D}} a(\omega,x). \tag{4.2}$$

We make the following assumptions on the input random field a and on the source term f :

A1. $a_{\min} \geq 0$ almost surely and $1/a_{\min} \in L^p(\Omega)$, for all $p \in (0,\infty)$.

A2. $a \in L^p(\Omega, C^t(\overline{D}))$, for some $0 < t \leq 1$ and for all $p \in (0,\infty)$.

A3. $f \in L^{p_*}(\Omega, H^{t-1}(D))$, for some $p_* \in (0,\infty]$.

The Hölder continuity of the trajectories of a in Assumption A2 implies that both quantities in (4.2) are well defined, that $a_{\max} \in L^p(\Omega)$ and (together with Assumption A1) that $a_{\min}(\omega) > 0$ and $a_{\max}(\omega) < \infty$, for almost all $\omega \in \Omega$. We will here not make the assumption that we can bound $a_{\min}(\omega)$ away from zero and $a_{\max}(\omega)$ away from infinity, uniformly in ω, as this is not true for log-normal fields a, for example. Many authors work with such assumptions of uniform ellipticity and boundedness. We shall instead work with the quantities $a_{\min}(\omega)$ and $a_{\max}(\omega)$ directly.

Note also that our assumptions on the (spatial) regularity of the coefficient function a are significantly weaker than what is usually assumed in the literature. Most other analyses of this problem assume at least that a has a gradient that lies in $L^\infty(D)$. As we will see below, we could even weaken Assumptions A1 and A2 and only assume that $\|a\|_{C^t(\overline{D})}$ and $1/a_{\min}$ have a finite number of bounded moments, i.e. $0 < p \leq p_a$, for some fixed $p_a > 0$, but in order not to complicate the presentation we did not choose to do this.

As usual in finite element methods, we will study the PDE (4.1) in weak (or variational) form, for fixed $\omega \in \Omega$. This is not possible uniformly in Ω, but almost surely. In the following we will not explicitly write this each time. With $f(\omega,\cdot) \in H^{t-1}(D)$ and $0 < a_{\min}(\omega) \leq a(\omega,x) \leq a_{\max}(\omega) < \infty$, for all $x \in D$, the variational formulation of (4.1), parametrised by $\omega \in \Omega$, is

$$b_\omega\big(u(\omega,\cdot), v\big) = L_\omega(v)\,, \quad \text{for all} \quad v \in H^1_0(D), \tag{4.3}$$

where the bilinear form b_ω and the linear functional L_ω (both parametrised by $\omega \in \Omega$) are defined as usual, for all $u, v \in H^1_0(D)$, by

$$b_\omega(u,v) := \int_D a(\omega,x)\nabla u(x) \cdot \nabla v(x)\,\mathrm{d}x \quad \text{and} \quad L_\omega(v) := \langle f(\omega,\cdot), v\rangle_{H^{t-1}(D),H^{1-t}_0(D)}\,. \tag{4.4}$$

We say that for any $\omega \in \Omega$, $u(\omega,\cdot)$ is a weak solution of (4.1) iff $u(\omega,\cdot) \in H^1_0(D)$ and satisfies (4.3). The following result is classical. It is based on the Lax-Milgram Lemma [39].

Lemma 4.2.1. *For almost all $\omega \in \Omega$, the bilinear form $b_\omega(u,v)$ is bounded and coercive in $H^1_0(D)$ with respect to $|\cdot|_{H^1(D)}$, with constants $a_{\max}(\omega)$ and $a_{\min}(\omega)$, respectively. Moreover, there exists a unique solution $u(\omega,\cdot) \in H^1_0(D)$ to the variational problem (4.3) and*

$$|u(\omega,\cdot)|_{H^1(D)} \lesssim \frac{\|f(\omega,\cdot)\|_{H^{t-1}(D)}}{a_{\min}(\omega)}.$$

The following proposition is a direct consequence of Lemma 4.2.1.

Theorem 4.2.2. *The weak solution u of (4.1) is unique and belongs to $L^p(\Omega, H^1_0(D))$, for all $p < p_*$.*

Proof. First note that $u : \Omega \to H^1_0(D)$ is measurable, since u is a continuous function of a. The result then follows directly from Lemma 4.2.1, Assumptions A1 and A3 and from the Hölder inequality. □

However, as usual in finite element methods we will require more (spatial) regularity of the solution u to be able to show convergence. We will come back to this in Section 4.3.1. First let us look at some typical examples of input random fields used in applications.

4.2.3 Log-normal Random Fields

A coefficient $a(\omega, x)$ of particular interest in applications of (4.1) is a log-normal random field, where $a(\omega, x) = \exp[g(\omega, x)]$ with $g : \Omega \times \overline{D} \to \mathbb{R}$ denoting a Gaussian field. We consider only mean zero homogeneous Gaussian fields with Lipschitz continuous covariance kernel

$$C(x, y) = \mathbb{E}\Big[(g(\omega, x) - \mathbb{E}[g(\omega, x)])(g(\omega, y) - \mathbb{E}[g(\omega, y)])\Big] = k\big(\|x - y\|\big), \quad \text{for some } k \in \mathcal{C}^{0,1}(\mathbb{R}^+) \tag{4.5}$$

and for some norm $\|\cdot\|$ in $\mathbb{R}^d$. In particular, this may be the usual modulus $\|x\| = |x| := (x^T x)^{1/2}$ or the 1-norm $\|x\| = \|x\|_1 := \sum_{i=1}^d |x_i|$. A typical example of a covariance function used in practice that is only Lipschitz continuous is the exponential covariance function given by

$$k(r) = \sigma^2 \exp(-r/\lambda), \tag{4.6}$$

for some parameters σ^2 (*variance*) and λ (*correlation length*).

With this type of covariance function, it follows from Kolmogorov's Theorem [13] that, for all $t < 1/2$, the trajectories of g belong to $\mathcal{C}^t(\overline{D})$ almost surely. More precisely, Kolmogorov's Theorem ensures the existence of a version $\tilde{g}$ of g (i.e. for any $x \in D$, we have $g(\cdot, x) = \tilde{g}(\cdot, x)$ almost surely) such that $\tilde{g}(\omega, \cdot) \in \mathcal{C}^t(\overline{D})$, for almost all $\omega \in \Omega$. In particular, we have for almost all ω, that $g(\omega, \cdot) = \tilde{g}(\omega, \cdot)$ almost everywhere. We will identify g with $\tilde{g}$ in what follows.

Built on the Hölder continuity of the trajectories of g and using Fernique's Theorem [13], it was shown in [10] that Assumption A1 holds and that $a \in L^p(\Omega, C^0(\overline{D}))$, for all $p \in (0, \infty)$.

Lemma 4.2.3. *Let g be Gaussian with covariance (4.5). Then the trajectories of the log-normal field $a = \exp g$ belong to $\mathcal{C}^t(\overline{D})$ almost surely, for all $t < r$, and*

$$\|a(\omega)\|_{\mathcal{C}^t} \;\le\; \Big(1 + 2\,|g(\omega)|_{\mathcal{C}^t}\Big)\, a_{\max}(\omega)\,.$$

Proof. Fix $\omega \in \Omega$ and $t < 1/2$. Since the trajectories of g belong to $\mathcal{C}^t(\overline{D})$ almost surely, we have

$$|e^{g(\omega,x)} - e^{g(\omega,y)}| \;\le\; |g(\omega, x) - g(\omega, y)| \left(e^{g(\omega,x)} + e^{g(\omega,y)}\right) \;\le\; 2\, a_{\max}(\omega)\, |g(\omega)|_{\mathcal{C}^t}\, |x - y|^t.$$

for any $x, y \in D$. Now, $a_{\max}(\omega)\, |g(\omega)|_{\mathcal{C}^t} < \infty$ almost surely, and so the result follows by taking the supremum over all $x, y \in D$. □

Lemma 4.2.3 can in fact be generalised from the exponential function to any smooth function of g.

Proposition 4.2.4. *Let g be a mean zero Gaussian field with covariance (4.5). Then Assumptions A1–A2 are satisfied for the log-normal field $a = \exp g$ with any $t < \frac{1}{2}$.*

Proof. Clearly by definition $a_{\min} \ge 0$. The proof that $1/a_{\min} \in L^p(\Omega)$, for all $p \in (0, \infty)$, is based on an application of Fernique's Theorem [13] and can be found in [10, Proposition 2.2]. To prove Assumption A2 note that, for all $t < 1/2$ and $p \in (0, \infty)$, $g \in L^p(\Omega, \mathcal{C}^t(\overline{D}))$ (cf. [10, Proposition 3.5]) and $a_{\max} \in L^p(\Omega)$ (cf. [10, Proposition 2.2]). Thus the result follows from Lemma 4.2.3 and an application of Hölder's inequality. □

Smoother covariance functions, such as the Gaussian covariance kernel

$$k(r) = \sigma^2 \exp(-r^2/\lambda^2), \tag{4.7}$$

which is analytic on $\overline{D} \times \overline{D}$, or more generally the covariance functions in the Matérn class with $\nu > 1$, all lead to $g \in C^1(\overline{D})$ and thus Assumption A2 is satisfied for all $t \le 1$.

Remark 4.2.5. The results in this section can be extended to log-normal random fields for which the underlying Gaussian field $g(\omega, x)$ does not have mean zero, under the assumption that this mean is sufficiently regular. Adding a mean $c(x)$ to g, we have $a(\omega, x) = \exp[c(x)] \exp[g(\omega, x)]$, and assumptions A1-A2 can still be satisfied. In particular, the assumptions still hold if $c(x) \in \mathcal{C}^{1/2}(\overline{D})$.

4.2.4 Truncated Karhunen-Loève Expansions

A starting point for many numerical schemes for PDEs with random coefficients is the approximation of the random field $a(\omega, x)$ as a function of a finite number of random variables, $a(\omega, x) \approx a(\xi_1(\omega), ..., \xi_K(\omega), x)$. This is true, for example, for the stochastic collocation method discussed in Section 4.3.4. Sampling methods, such as Monte–Carlo type methods discussed in Section 4.4, do not rely on such a finite–dimensional approximation as such, but may make use of such approximations as a way of producing samples of the input random field.

A popular choice to achieve good approximations of this kind for log-normal random fields is the truncated Karhunen-Loève (KL) expansion. For the random field g (as defined in Section 4.2.2), the KL–expansion is an expansion in terms of a countable set of independent, standard Gaussian random variables $\{\xi_n\}_{n\in\mathbb{N}}$. It is given by

$$g(\omega, x) = \sum_{n=1}^{\infty} \sqrt{\theta_n} b_n(x) \xi_n(\omega),$$

where $\{\theta_n\}_{n\in\mathbb{N}}$ are the eigenvalues and $\{b_n\}_{n\in\mathbb{N}}$ the corresponding normalised eigenfunctions of the covariance operator with kernel function $C(x, y)$ defined in (4.5). For more details on its derivation and properties, see e.g. [31]. We will here only mention that the eigenvalues $\{\theta_n\}_{n\in\mathbb{N}}$ are all non–negative with $\sum_{n\geq 0} \theta_n < +\infty$.

We shall write $a(\omega, x)$ as

$$a(\omega, x) = \exp\left[\sum_{n=1}^{\infty} \sqrt{\theta_n} b_n(x) \xi_n(\omega)\right],$$

and denote the random fields resulting from truncated expansions by

$$g_K(\omega, x) := \sum_{n=1}^{K} \sqrt{\theta_n} b_n(x) \xi_n(\omega) \quad \text{and } a_K(\omega, x) := \exp\left[\sum_{n=1}^{K} \sqrt{\theta_n} b_n(x) \xi_n(\omega)\right], \quad \text{for some } K \in \mathbb{N}.$$

The advantage of writing $a(\omega, x)$ in this way is that it gives an expression for $a(\omega, x)$ in terms of independent, standard Gaussian random variables.

Finally, we denote by u_K the solution to our model problem with the coefficient replaced by its truncated approximation,

$$-\nabla \cdot (a_K(\omega, x) \nabla u_K(\omega, x)) = f(\omega, x), \tag{4.8}$$

subject to homogeneous Dirichlet boundary conditions $u_K = 0$ on ∂D.

Under certain additional assumptions on the eigenvalues and eigenfunctions of the covariance operator, it is possible to bound the truncation error $\|u - u_K\|_{L^p(\Omega, H_0^1(D))}$, for all $p < p_*$, as shown in [10]. We make the following assumptions on $(\theta_n, b_n)_{n\geq 0}$:

B1. The eigenfunctions are continuously differentiable, i.e. $b_n \in \mathcal{C}^1(\overline{D})$ for any $n \geq 0$.

B2. We have

$$\sum_{n\geq 0} \theta_n \|b_n\|^2_{L^\infty(D)} < +\infty.$$

B3. There exists an $r \in (0, 1)$ such that,

$$\sum_{n\geq 0} \theta_n \|b_n\|^{2(1-r)}_{L^\infty(D)} \|\nabla b_n\|^{2r}_{L^\infty(D)} < +\infty.$$

Proposition 4.2.6. *Let Assumptions B1–B2–B3 be satisfied, for some $r \in (0, 1)$. Then Assumptions A1–A2 are satisfied also for the truncated KL–expansion a_K of a, for any $K \in \mathbb{N}$ and $t < r$. Moreover, $\|a_K\|_{L^p(\Omega, \mathcal{C}^t(\bar{D}))}$ and $\|1/a_K^{min}\|_{L^p(\Omega)}$ can be bounded independently of K. If, moreover, Assumptions B1–B2–B3 are satisfied for $\frac{\partial b_n}{\partial x_i}$ instead of b_n, for all $i = 1, \ldots, d$, then $\|a_K\|_{L^p(\Omega, \mathcal{C}^1(\bar{D}))}$ can be bounded independently of K.*

Proof. This is essentially [10, Propositions 3.7 and 3.8]. The Hölder continuity $a_K \in L^p(\Omega, C^t(\overline{D})$, for all $p \in (0, \infty)$, can be deduced as in Lemma 4.2.3 from the almost sure Hölder continuity of the trajectories of g_K proved in [10, Proposition 3.5]. □

Let us recall the main result on the strong convergence of u to u_K from [10, Theorem 4.2].

Theorem 4.2.7. *Let Assumptions B1–B2–B3 be satisfied, for some $r \in (0,1)$. Then, u_K converges, for all $p < p_*$, to $u \in L^p(\Omega, H_0^1(D))$. Moreover, for any $s \in [0,1]$, we have*

$$\|u - u_K\|_{L^p(\Omega,H_0^1(D))} \lesssim \underbrace{\Big(\sum_{n>K} \theta_n \|b_n\|_{L^\infty(D)}^{2(1-r)} \|\nabla b_n\|_{L^\infty(D)}^{2r} \Big)^{1/2}}_{=: C_{4.2.7}(K)} \|f\|_{L^p(\Omega,H^{s-1}(D))}$$

The hidden constant depends only on D, r, p. Similarly $\|u - u_K\|_{L^p(\Omega,L^2(D))} \lesssim C_{4.2.7} \|f\|_{L^p(\Omega,H^{s-1}(D))}$.

Assumptions B1–B2–B3 are fulfilled, among other cases, for the analytic covariance function (4.7) as well as for the exponential covariance function (4.6) with 1-norm $\|x\| = \sum_{i=1}^d |x_i|$ in (4.5), since then the eigenvalues and eigenvectors can be computed explicitly and we have explicit decay rates for the KL–eigenvalues. For details see [10, Section 7]. In the latter case, on non-rectangular domains D, we need to use a KL-expansion on a bounding box containing D to get again explicit formulae for the eigenvalues and eigenvectors. Strictly speaking this is not a KL–expansion on D.

Proposition 4.2.8. *We have the following bound on the constant in Theorem 4.2.7 :*

$$C_{4.2.7}(K) \lesssim \begin{cases} K^{\frac{\rho-1}{2}}, & \text{for the 1-norm exponential covariance kernel,} \\ K^{\frac{d-1}{2d}} \exp\big(-c_1 K^{1/d}\big), & \text{for the Gaussian covariance kernel,} \end{cases}$$

for some constant $c_1 > 0$ and for any $0 < \rho < 1$. The hidden constants depend only on D, ρ, p.

4.3 Finite Element Error Analysis

Let us now come to the central part of this paper. To carry out a finite element error analysis for (4.1) under Assumptions A1-A3, we will require first of all regularity results for trajectories of the weak solution u. However, since we did not assume uniform ellipticity or boundedness of $b_\omega(\cdot,\cdot)$, it is essential that we track exactly, how the constants in the regularity estimates depend on the input random field a. We were unable to find such a result in the literature, and we will therefore in Section 4.3.1 reiterate the steps in the usual regularity proof for the solution $u(\omega,\cdot)$ of (4.3) following the proof in Hackbusch [39], but highlighting explicitly where and how constants depend on a. A detailed proof is given in Appendix 4.7. We will need to assume that $D \subset \mathbb{R}^d$ is a $\mathcal{C}^2$ bounded domain for this proof. Regularity proofs that require only Lipschitz continuity for the boundary of D do exist (cf. Hackbusch [39, Theorems 9.1.21 and 9.1.22]), but we did not consider it instructive to complicate the presentation even further. We do not see any obstacles in also getting an explicit dependence of the constants in the case of Lipschitz continuous boundaries.

Having established the regularity of trajectories of u, we then carry out a classical finite element error analysis for each trajectory in Section 4.3.2 and deduce error estimates for the moments of the error in u. Since we only have a regularity result for $\mathcal{C}^2$ bounded domains and since the integrals appearing in $b_\omega(\cdot,\cdot)$ can in general only be approximated by quadrature, we will also need to address variational crimes, such as the approximation of a $\mathcal{C}^2$ bounded domain by a polygonal domain as well as quadrature errors (cf. Section 4.3.3).

Let us assume for the rest of this section that $D \subset \mathbb{R}^d$ is a $\mathcal{C}^2$ bounded domain.

4.3.1 Regularity of the Solution

Proposition 4.3.1. *Let Assumptions A1-A3 hold with $0 < t < 1$. Then, $u(\omega,\cdot) \in H^{1+s}(D)$, for all $0 < s < t$ except $s = 1/2$ almost surely in $\omega \in \Omega$, and*

$$\|u(\omega,\cdot)\|_{H^{1+s}(D)} \lesssim C_{4.3.1}(\omega) \|f(\omega,\cdot)\|_{H^{s-1}(D)}, \quad \text{where} \quad C_{4.3.1}(\omega) := \frac{a_{\max}(\omega)\|a(\omega,\cdot)\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}(\omega)^3}.$$

If the assumptions hold with $t = 1$, then $u(\omega,\cdot) \in H^2(D)$ and $\|u(\omega,\cdot)\|_{H^2(D)} \lesssim C_{4.3.1}(\omega) \|f(\omega,\cdot)\|_{L^2(D)}$.

We give here only the main elements of the proof and consider only the case $t < 1$ in detail. It follows the proof of [39, Theorem 9.1.16] and consists in three main steps. We formulate the first two steps as separate lemmas and then give the final step following the lemmas. We fix $\omega \in \Omega$ and to simplify the notation we will not specify the dependence on ω anywhere in the proof.

In the first step of the proof we take $D = \mathbb{R}^d$ and establish the regularity of a slightly more general elliptic PDE with tensor-valued coefficients.

Lemma 4.3.2. *Let $0 < t < 1$ and $D = \mathbb{R}^d$, and let $A = (A_{ij})_{i,j=1}^d \in S_d(\mathbb{R})$ be a symmetric, uniformly positive definite $n \times n$ matrix-valued function from D to $\mathbb{R}^{n\times n}$, i.e. there exists $A_{\min} > 0$ such that $A(x)\xi\cdot\xi \geq A_{\min}|\xi|^2$ uniformly in $x \in \overline{D}$ and $\xi \in \mathbb{R}^d$, and let $A_{ij} \in \mathcal{C}^t(\overline{D})$, for all $i, j = 1, \ldots, d$. Consider*

$$-\mathrm{div}(A(x)\nabla w(x)) = F(x) \qquad \text{in} \quad D \tag{4.9}$$

with $F \in H^{s-1}(\mathbb{R}^d)$, for some $0 < s < t$. Any weak solution $w \in H^1(\mathbb{R}^d)$ of (4.9) is in $H^{1+s}(\mathbb{R}^d)$ and

$$\|w\|_{H^{1+s}(\mathbb{R}^d)} \lesssim \frac{1}{A_{\min}}\Big(|A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)}\Big) \;+\; \|w\|_{H^1(\mathbb{R}^d)},$$

where $|A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))}$ is the Hölder seminorm on $S_d(\mathbb{R})$ using a suitable matrix norm.

Proof. This is essentially [39, Theorem 9.1.8] with the dependence on A made explicit, and it can be proved using the representation of the norm on $H^{1+s}(\mathbb{R}^d)$ via Fourier coefficients, as well as a fractional difference operator R_h^i, $i = 1, \ldots, d$, on a Cartesian mesh with mesh size $h > 0$ (similar to the classical Nirenberg translation method for proving H^2 regularity).

It is shown in the proof of [39, Theorem 9.1.8] that $\sum_{i=1}^d \|(R_h^i)^* w\|_{H^1(\mathbb{R}^d)}$ is an upper bound for $\|w\|_{H^{1+s}(\mathbb{R}^d)}$ and that it can itself be bounded from above in terms of $\|w\|_{H^1(\mathbb{R}^d)}$ and $\|F\|_{H^{s-1}(\mathbb{R}^d)}$ using the weak form (4.3) of our problem. The dependence of this upper bound on $A_{\min}$ stems from the fact that in order to use (4.3), we have to switch from $|(R_h^i)^* w|_{H^1(\mathbb{R}^d)}$ to the energy norm $\int_{\mathbb{R}^d} A(x)|\nabla(R_h^i)^* w|^2\,\mathrm{d}x$. The dependence on $|A_{ij}|_{\mathcal{C}^t(\mathbb{R}^d)}$ comes from bounding the differences of A_{ij} at two consecutive grid points in the translation step.

For a definition of R_h^i and more details see [39, Theorem 9.1.8] or Section 4.7.1 in the appendix. □

The second step consists in treating the case where $D = \mathbb{R}^d_+ := \{y = (y_1, ..., y_d) : y_d > 0\}$.

Lemma 4.3.3. *Let $0 < t < 1$ and $D = \mathbb{R}^d_+$, and let $A : D \to S_d(\mathbb{R})$ be as in Lemma 4.3.2. Consider now (4.9) on $D = \mathbb{R}^d_+$ subject to $w = 0$ on ∂D with $F \in H^{s-1}(\mathbb{R}^d_+)$, for some $0 < s < t$, $s \neq 1/2$. Then any weak solution $w \in H^1(\mathbb{R}^d_+)$ of this problem is in $H^{1+s}(\mathbb{R}^d_+)$ and*

$$\|w\|_{H^{1+s}(\mathbb{R}^d_+)} \;\lesssim\; \frac{A_{\max}}{A_{\min}^2}\Big(|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d_+)} + \|F\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) \;+\; \frac{A_{\max}}{A_{\min}}\,\|w\|_{H^1(\mathbb{R}^d_+)}\,.$$

where $A_{\max} := \|A\|_{\mathcal{C}^0(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}$.

Proof. This is essentially [39, Theorem 9.1.11] with the dependence on A made explicit. It uses the fact that for any $s > 0$, $s \neq 1/2$, the norm

$$|||v|||_s := \Big(\|v\|^2_{L^2(\mathbb{R}^d_+)} + \sum_{i=1}^d \Big\|\tfrac{\partial v}{\partial x_i}\Big\|^2_{H^{s-1}(\mathbb{R}^d_+)}\Big)^{1/2}$$

is equivalent to the usual norm on $H^s(\mathbb{R}^d_+)$. Then using the same approach as in the proof of Lemma 4.3.2, we can establish that, for $1 \leq j \leq d-1$,

$$\Big\|\frac{\partial w}{\partial x_j}\Big\|_{H^s(\mathbb{R}^d_+)} \;\lesssim\; \frac{1}{A_{\min}}\Big(|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d_+)} + \|F\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) + \|w\|_{L^2(\mathbb{R}^d_+)}\,. \tag{4.10}$$

To establish a similar bound for $j = d$ is technical. We use (4.9) and the following inequality, for any $D \subset \mathbb{R}^d$ and $0 < s < t < 1$:

$$\|bv\|_{H^s(D)} \;\lesssim\; |b|_{\mathcal{C}^t(\overline{D})}\|v\|_{L^2(D)} \;+\; \|b\|_{\mathcal{C}^0(\overline{D})}\|v\|_{H^s(D)}, \quad \text{for all } b \in \mathcal{C}^t(\overline{D}),\ v \in H^s(D). \tag{4.11}$$

We use (4.11) to bound the H^s–norm of $A_{ij}\frac{\partial w}{\partial x_j}$, for $(i,j) \neq (d,d)$, and so by rearranging the weak form of (4.9) we can also bound the H^s–norm of $A_{dd}\frac{\partial w}{\partial x_d}$. This leads to the additional factor $A_{\max}$ in the bound. The final result can then be deduced by applying (4.11) once more, with $b = 1/A_{dd}$ and $v = A_{dd}\frac{\partial w}{\partial x_d}$, leading to an additional factor $1/A_{\min}$. For details see [39, Theorem 9.1.11] or Section 4.7.2 in the appendix. □

Proof of Proposition 4.3.1. We are now ready to prove Proposition 4.3.1 using Lemmas 4.3.2 and 4.3.3. The third and last step consists in using a covering of D by $m+1$ bounded regions $(D_i)_{0\leq i\leq m}$, such that

$$\overline{D}_0 \subset D, \quad \overline{D} \subset \bigcup_{i=0}^{m} D_i \quad \text{and} \quad \partial D = \bigcup_{i=1}^{m} (D_i \cap \partial D).$$

Using a (non-negative) partition of unity $\{\chi_i\}_{0\leq i\leq m} \subset \mathcal{C}^\infty(\mathbb{R}^d)$ subordinate to this cover, it is possible to reduce the proof to bounding $\|\chi_i u\|_{H^{1+s}(D)}$, for all $0 \leq i \leq m$.

For $i = 0$ this reduces to an application of Lemma 4.3.2 with w and F chosen to be extensions by 0 from D to $\mathbb{R}^d$ of $\chi_0 u$ and of $f\chi_0 + a\nabla u\cdot\nabla\chi_0 + \mathrm{div}(au\nabla\chi_0)$, respectively. The tensor A degenerates to $\overline{a}(x)I_d$, where $\overline{a}$ is a smooth extension of $a(x)$ on D_0 to $a_{\min}$ on $\mathbb{R}^d\backslash D$, and so $A_{\min} = a_{\min}$ and $|A|_{\mathcal{C}^t(\mathbb{R}^d,S_d(\mathbb{R}))} \leq |a|_{\mathcal{C}^t(D)}$.

For $1 \leq i \leq m$, the proof reduces to an application of Lemma 4.3.3. As for $i = 0$, we can see that $\chi_i u \in H^1_0(D \cap D_i)$ is the weak solution of the problem $-\mathrm{div}(a\nabla u_i) = f_i$ on $D \cap D_i$ with $f_i := f\chi_i + a\nabla u \cdot \nabla\chi_i + \mathrm{div}(au\nabla\chi_i)$. To be able to apply Lemma 4.3.3 to the weak form of this PDE, we define now a twice continuously differentiable bijection α_i (with α_i^{-1} also in $\mathcal{C}^2$) from D_i to the cylinder

$$Q_i := \{y = (y_1, \ldots, y_d) : |(y_1, \ldots, y_{d-1})| < 1 \ \text{ and } \ |y_d| < 1\},$$

such that $D_i \cap D$ is mapped to $Q_i \cap \mathbb{R}^d_+$, and $D_i \cap \partial D$ is mapped to $Q_i \cap \{y : y_d = 0\}$. We use α_i^{-1} to map all the functions defined above on $D_i \cap D$ to $Q_i \cap \mathbb{R}^d_+$, and then extend them suitably to functions on $\mathbb{R}^d_+$ to finally apply Lemma 4.3.3. The tensor A in this case is a genuine tensor depending on the mapping α_i. However, since ∂D was assumed to be $\mathcal{C}^2$, we get $A_{\min} \lesssim a_{\min}$, $A_{\max} \lesssim a_{\max}$ and $|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))} \lesssim |a|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+})}$, with hidden constants that only depend on α_i, α_i^{-1} and their Jacobians. For details see [39, Theorem 9.1.16] or Section 4.7.3 in the appendix. □

Using Proposition 4.3.1 and Assumptions A1-A3, we can now conclude on the regularity of u.

Theorem 4.3.4. *Let Assumptions A1-A3 hold with $0 < t \leq 1$. Then $u \in L^p(\Omega, H^{1+s}(D))$, for all $p < p_*$ and for all $0 < s < t$ except $s = 1/2$. If $t = 1$, then $u \in L^p(\Omega, H^2(D))$.*

Proof. Since $a_{\max}(\omega) \leq \|a(\omega,\cdot)\|_{\mathcal{C}^t(\overline{D})}$ and $H^{s-1}(D) \subset H^{t-1}(D)$, for all $s < t$, the result follows directly from Proposition 4.3.1 and Assumptions A1–A3 via Hölder's inequality. □

Remark 4.3.5. Note that in order to establish $u \in L^p(\Omega, H^{1+s}(D))$, for some fixed $p > 0$, it would have been sufficient to assume that the constant $C_{4.3.1}$ in Proposition 4.3.1 is in $L^q(\Omega)$ for $q = \frac{p_* p}{p_* - p}$. In the case $p_* = \infty$, $q = p$ is sufficient. This in turn implies that we can weaken Assumption A1 to $1/a_{\min} \in L^q(\Omega)$ with $q > 3p$, or Assumption A2 to $a \in L^q(\Omega, C^t(\overline{D}))$ with $q > 2p$, or both assumptions to L^q with $q > 5p$.

However, in the case of a log-normal field a and $p_* = \infty$, we do have bounds on all moments $p \in (0,\infty)$, but in general we only have the limited spatial regularity of $1 + s < 3/2$.

4.3.2 Finite Element Approximation

We consider finite element approximations of our model problem (4.1) using standard, continuous, piecewise linear finite elements. The aim is to derive estimates of the finite element error in the $L^p(\Omega, H^1_0(D))$ and $L^p(\Omega, L^2(D))$ norms. To remain completely rigorous, we keep the assumption that boundary of D is $\mathcal{C}^2$, so that we can apply the explicit regularity results from the previous section. However, this means that we will have to approximate our domain D by polygonal domains D_h in dimensions $d \geq 2$.

We denote by $\{\mathcal{T}_h\}_{h>0}$ a shape-regular family of simplicial triangulations of the domain D, parametrised by its mesh width $h := \max_{\tau\in\mathcal{T}_h} \mathrm{diam}(\tau)$, such that, for any $h > 0$,

– $\overline{D} \subset \bigcup_{\tau\in\mathcal{T}_h} \tau$, i.e. the triangulation covers all of $\overline{D}$, and

– the vertices $x_1^\tau, \dots, x_{d+1}^\tau$ of any $\tau \in \mathcal{T}_h$ lie either all in $\overline{D}$ or all in $\mathbb{R}^d \backslash D$.

Let $\overline{D}_h$ denote the union of all simplices that are interior to $\overline{D}$ and D_h its interior, so that $D_h \subset D$.

Associated with each triangulation $\mathcal{T}_h$ we define the space

$$V_h := \Big\{ v_h \in C(\overline{D}) : v_h|_\tau \text{ linear, for all } \tau \in \mathcal{T}_h \text{ with } \tau \subset \overline{D}_h\,, \quad \text{and} \quad v_h|_{\overline{D}\backslash D_h} = 0 \Big\} \tag{4.12}$$

of continuous, piecewise linear functions on D_h that vanish on the boundary of D_h and in $D\backslash D_h$. Let us recall the following standard interpolation result (see e.g. [8, Section 4.4]).

Lemma 4.3.6. *Let $v \in H^{1+s}(D_h)$, for some $0 < s \leq 1$. Then*

$$\inf_{v_h \in V_h} |v - v_h|_{H^1(D_h)} \lesssim \|v\|_{H^{1+s}(D_h)}\, h^s. \tag{4.13}$$

The hidden constant is independent of h and v.

This can easily be extended to an interpolation result for functions $v \in H^{1+s}(D) \cap H_0^1(D)$, by estimating the residual over $D\backslash D_h$. However, when D is not convex it requires local mesh refinement in the vicinity of any non-convex parts of the boundary. We make the following assumption on $\mathcal{T}_h$:

A4 For all $\tau \in \mathcal{T}_h$ with $\tau \cap D_h = \emptyset$ and $x_1^\tau, \dots, x_{d+1}^\tau \in \overline{D}$, we assume $\operatorname{diam}(\tau) \lesssim h^2$.

Lemma 4.3.7. *Let $v \in H^{1+s}(D) \cap H_0^1(D)$, for some $0 < s \leq 1$, and let Assumption A4 hold. Then*

$$\inf_{v_h \in V_h} |v - v_h|_{H^1(D)} \lesssim \|v\|_{H^{1+s}(D)}\, h^s. \tag{4.14}$$

Proof. This result is classical (for parts of the proof see [39, Section 8.6] or [77]). Set $D_\delta := D\backslash\overline{D}_h$ where δ denotes the maximum width of D_δ, and let first $s = 1$. Since $v_h = 0$ on D_δ it suffices to show that

$$|v|_{H^1(D_\delta)} \lesssim \|v\|_{H^2(D)}\, h\,. \tag{4.15}$$

The result then follows for $s = 1$ with Lemma 4.3.6. The result for $s < 1$ follows by interpolation, since trivially, $|v|_{H^1(D_\delta)} \leq \|v\|_{H^1(D)}$,

To show (4.15), let $w \in H^1(D)$. Using a trace result we get

$$\|w\|_{L^2(D_\delta)} \leq \|w\|_{L^2(S_\delta)} \lesssim \delta^{1/2} \|w\|_{H^1(D)},$$

where $S_\delta = \{x \in D : \operatorname{dist}(x, \partial D) \leq \delta\} \subset D$ is the boundary layer of width δ. It follows from Assumption A4 that $\operatorname{diam}(\tau) \lesssim h^2$ wherever the boundary is not convex. In regions where D is convex it follows from the smoothness assumption on ∂D that the width of D_δ is $\mathcal{O}(h^2)$. Hence $\delta \lesssim h^2$, which completes the proof of (4.15). □

Now, for almost all $\omega \in \Omega$, the finite element approximation to (4.3), denoted by $u_h(\omega, \cdot)$, is the unique function in V_h that satisfies

$$b_\omega(u_h(\omega, \cdot), v_h) = L_\omega(v_h), \qquad \text{for all} \quad v_h \in V_h. \tag{4.16}$$

Since $b_\omega(\cdot, \cdot)$ is coercive and bounded in $H_0^1(D)$ (cf. Lemma 4.2.1), we have

$$|u_h(\omega, \cdot)|_{H^1(D)} \lesssim \|f(\omega, \cdot)\|_{H^{t-1}(D)} / a_{\min}(\omega). \tag{4.17}$$

as well as the following classical quasi optimality result (cf. [8, 39]).

Lemma 4.3.8 (Cea's Lemma). *Let Assumptions A1–A3 hold. Then, for almost all $\omega \in \Omega$,*

$$|u(\omega, \cdot) - u_h(\omega, \cdot)|_{H^1(D)} \leq C_{4.3.8}(\omega) \inf_{v_h \in V_h} |u(\omega, \cdot) - v_h|_{H^1(D)}, \quad \textit{where} \quad C_{4.3.8}(\omega) := \left(\frac{a_{\max}(\omega)}{a_{\min}(\omega)} \right)^{1/2}.$$

Combining this with the interpolation result above we get the following error estimates.

Theorem 4.3.9. *Let Assumptions A1–A4 hold, for some $0 < t < 1$ and $p_* \in (0, \infty]$. Then, for all $p < p_*$, $s < t$ with $s \neq 1/2$ and $h > 0$, we have*

$$\|u - u_h\|_{L^p(\Omega, H^1_0(D))} \lesssim C_{4.3.9} \, \|f\|_{L^{p_*}(\Omega, H^{s-1}(D))} \, h^s, \quad \text{where} \quad C_{4.3.9} := \left\| \frac{a_{\max}^{3/2} \|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}^{7/2}} \right\|_{L^q(\Omega)}$$

with $q = \frac{p_* \, p}{p_* - p}$. If Assumptions A2 and A3 hold with $t = 1$, then

$$\|u - u_h\|_{L^p(\Omega, H^1_0(D))} \lesssim C_{4.3.9} \, \|f\|_{L^{p_*}(\Omega, L^2(D))} \, h \, .$$

Proof. Let $0 < t < 1$ in Assumptions A2 and A3. It follows from Proposition 4.3.1 and Lemmas 4.3.7 and 4.3.8 that, for almost all $\omega \in \Omega$,

$$|u(\omega, \cdot) - u_h(\omega, \cdot)|_{H^1(D)} \lesssim C_{4.3.1}(\omega) \, C_{4.3.8}(\omega) \, \|f(\omega, \cdot)\|_{H^{s-1}(D)} \, h^s. \tag{4.18}$$

The result now follows by Hölder's inequality. The proof for $t = 1$ is analogous. □

The usual duality (or Aubin–Nitsche) "trick" leads to a bound on the L^2–error.

Corollary 4.3.10. *Let Assumptions A1–A4 hold, for some $0 < t < 1$ and $p_* \in (0, \infty]$. Then, for all $p < p_*$, $s < t$ with $s \neq 1/2$ and $h > 0$, we have*

$$\|u - u_h\|_{L^p(\Omega, L^2(D))} \lesssim C_{4.3.10} \, \|f\|_{L^{p_*}(\Omega, H^{s-1}(D))} \, h^{2s}, \quad \text{where} \quad C_{4.3.10} := \left\| \frac{a_{\max}^{7/2} \|a\|^2_{\mathcal{C}^t(\overline{D})}}{a_{\min}^{13/2}} \right\|_{L^q(\Omega)}$$

with $q = \frac{p_* \, p}{p_* - p}$. If Assumptions A2 and A3 hold with $t = 1$, then

$$\|u - u_h\|_{L^p(\Omega, L^2(D))} \lesssim C_{4.3.10} \, \|f\|_{L^{p_*}(\Omega, L^2(D))} \, h^2.$$

Proof. We will use a duality argument. For almost all $\omega \in \Omega$, let $e_\omega := u(\omega, \cdot) - u_h(\omega, \cdot)$ and denote by w_ω the solution to the adjoint problem

$$b_\omega(v, w_\omega) = (e_\omega, v) \qquad \text{for all } v \in H^1_0(D),$$

which, by Proposition 4.3.1 with $f(\omega, \cdot) = e_\omega$, is also in $H^{1+s}(D)$. By Galerkin orthogonality

$$\|e_\omega\|^2_{L^2(D)} = (e_\omega, e_\omega)_{L^2(D)} = b_\omega(e_\omega, w_\omega - z_h), \quad \text{for any} \quad z_h \in V_h.$$

Using the boundedness of $b_\omega(\cdot, \cdot)$ and Lemma 4.3.7, we then get

$$\|e_\omega\|^2_{L^2(D)} \lesssim a_{\max}(\omega) \, |u(\omega, \cdot) - u_h(\omega, \cdot)|_{H^1(D)} \, \|w_\omega\|_{H^{1+s}(D)} \, h^s.$$

Now it follows from Proposition 4.3.1 and Theorem 4.3.9 that

$$\|e_\omega\|^2_{L^2(D)} \leq a_{\max}(\omega) \, C_{4.3.1}(\omega)^2 \, C_{4.3.8}(\omega) \, \|f(\omega, \cdot)\|_{H^{s-1}(D)} \|e_\omega\|_{L^2(D)} \, h^{2s}.$$

Dividing by $\|e_\omega\|_{L^2(D)}$, the result follows again by an application of Hölder's inequality. □

Remark 4.3.11. As usual, the analysis can be extended in a straightforward way to other boundary conditions, to tensor coefficients, or to more complicated PDEs including low–order terms, provided the boundary data and the coefficients are sufficiently regular, see [39] for details.

4.3.3 Quadrature Error

The integrals appearing in the bilinear form

$$b_\omega(w_h, v_h) \;=\; \sum_{\tau\in\mathcal{T}_h:\,\tau\in\overline{D}_h} \int_\tau a(\omega,x)\nabla w_h\cdot\nabla v_h\,\mathrm{d}x$$

and in the linear functional $L_\omega(v_h)$ involve realisations of random fields. It will in general be impossible to evaluate these integrals exactly, and so we will use quadrature instead. We will only explicitly analyse the quadrature error in b_ω, but the quadrature error in approximating $L_\omega(v_h)$ can be analysed analogously.

In our analysis, we use the midpoint rule, approximating the integrand by its value at the midpoint x_τ of each simplex $\tau\in\mathcal{T}_h$. The trapezoidal rule which we use in the numerical section can be analysed analogously. Let us denote the resulting bilinear form that approximates b_ω on the grid $\mathcal{T}_h$ by

$$\tilde{b}_\omega(w_h, v_h) = \sum_{\tau\in\mathcal{T}_h:\,\tau\in\overline{D}_h} a(\omega,x_\tau)\int_\tau \nabla w_h(x)\cdot\nabla v_h(x)\,\mathrm{d}x\,,$$

and let $\tilde{u}_h(\omega,\cdot)$ denote the corresponding solution to

$$\tilde{b}_\omega(\tilde{u}_h(\omega,\cdot), v_h) = L_\omega(v_h)\,,\quad \text{for all}\quad v_h\in V_h.$$

Clearly the bilinear form $\tilde{b}_\omega$ is bounded and coercive, with the same constants as the exact bilinear form b_ω and so we can apply the following classical result [11] (with explicit dependence of the bound on the coefficients).

Lemma 4.3.12 (First Strang Lemma)**.** *Let Assumptions A1-A3 hold. Then, for almost all $\omega\in\Omega$,*

$$|u(\omega,\cdot)-\tilde{u}_h(\omega,\cdot)|_{H^1(D)} \le \inf_{v_h\in V_h}\left\{\left(1+\frac{a_{\max}(\omega)}{a_{\min}(\omega)}\right)|u(\omega,\cdot)-v_h|_{H^1(D)} + \frac{1}{a_{\min}(\omega)}\sup_{w_h\in V_h}\frac{|b_\omega(v_h,w_h)-\tilde{b}_\omega(v_h,w_h)|}{|w_h|_{H^1(D)}}\right\}.$$

Proposition 4.3.13. *Let Assumptions A1–A4 hold, for some $0<t<1$ and $p_*\in(0,\infty]$. Then, for all $p<p_*$, $s<t$ with $s\neq 1/2$ and $0<h<1$, we have*

$$\|u-\tilde{u}_h\|_{L^p(\Omega,H^1_0(D))} \;\lesssim\; C_{4.3.13}\,\|f\|_{L^{p_*}(\Omega,H^{s-1}(D))}\,h^s,\quad \textit{where}\quad C_{4.3.13} := \left\|\frac{a_{\max}^{5/2}\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}^{9/2}}\right\|_{L^q(\Omega)}$$

with $q=\frac{p_*\,p}{p_*-p}$. If Assumptions A2 and A3 hold with $t=1$, then

$$\|u-\tilde{u}_h\|_{L^p(\Omega,H^1_0(D))} \;\lesssim\; C_{4.3.13}\,\|f\|_{L^{p_*}(\Omega,L^2(D))}\,h\,.$$

Proof. We first note that, for all $w_h\in V_h$,

$$\begin{aligned}\left|b_\omega(v_h,w_h)-\tilde{b}_\omega(v_h,w_h)\right| \;&=\; \left|\sum_{\tau\in\mathcal{T}_h}\int_\tau (a(\omega,x)-a(\omega,x_\tau))\,\nabla v_h\cdot\nabla w_h\,\mathrm{d}x\right|\\ &\le \sum_{\tau\in\mathcal{T}_h}\int_\tau \frac{|a(\omega,x)-a(\omega,x_\tau)|}{|x-x_\tau|^t}\,|x-x_\tau|^t\,|\nabla v_h(\omega)\cdot\nabla w_h|\,\mathrm{d}x\\ &\le |a(\omega)|_{\mathcal{C}^t(\overline{D})}\,h^t\,|v_h|_{H^1(D)}\,|w_h|_{H^1(D)}\,.\end{aligned}$$

Hence, it follows from Lemma 4.3.12 that, for almost all $\omega\in\Omega$,

$$|u(\omega,\cdot)-\tilde{u}_h(\omega,\cdot)|_{H^1(D)} \le \inf_{v_h\in V_h}\left\{\left(1+\frac{a_{\max}(\omega)}{a_{\min}(\omega)}\right)|u(\omega,\cdot)-v_h|_{H^1(D)} + h^t\,\frac{|a(\omega)|_{\mathcal{C}^t(\overline{D})}}{a_{\min}(\omega)}\,|v_h|_{H^1(D)}\right\}.$$

Let us now make the particular choice $v_h := u_h(\omega,\cdot) \in V_h$, i.e. the solution of (4.16). Then it follows from (4.17) and (4.18) and the fact that $h^t < h^s$, for any $s < t \leq 1$ and $h < 1$, that

$$|u(\omega,\cdot) - \tilde{u}_h(\omega,\cdot)|_{H^1(D)} \lesssim \left(\left(1 + \frac{a_{\max}(\omega)}{a_{\min}(\omega)}\right) C_{4.3.1}(\omega)\, C_{4.3.8}(\omega) \;+\; \frac{|a(\omega)|_{\mathcal{C}^t(\overline{D})}}{a_{\min}(\omega)^2} \right) \|f(\omega,\cdot)\|_{H^{s-1}(D)}\, h^s.$$

The result follows again via an application of Hölder's inequality. □

Remark 4.3.14. To recover the $\mathcal{O}(h^{2s})$ convergence for the L^2-error $\|u - \tilde{u}_h\|_{L^p(\Omega,L^2(D))}$ in the case of quadrature, we require additional regularity of the coefficient function a. If $a(\omega,\cdot)$ is at least $\mathcal{C}^{2s}$, then we can again obtain $\mathcal{O}(h^{2s})$ convergence even with quadrature, using duality as in Corollary 4.3.10. This is for example the case in the context of lognormal random fields with Gaussian covariance kernel, where $a(\omega,\cdot) \in \mathcal{C}^\infty(\overline{D})$. In the context of the exponential covariance kernel the L^2-convergence rate is always bounded by $\mathcal{O}(h^{1/2-\delta})$, due to the lack of regularity in a.

4.3.4 Truncated Fields

Combining the results from Sections 4.2.4 and 4.3.2 we can also estimate the error in the case of truncated KL-expansions of log-normal coefficients which we will use in the numerical experiments in Section 4.5. These results give an improvement of the error estimate given in [10] for log-normal coefficients that do not admit full regularity, e.g. those with exponential covariance kernel. The novelty with respect to [10] is the improvement of the bound for the finite element error, leading to a uniform bound with respect to the number of terms in the KL–expansion. We will assume here that $f(\omega,\cdot) \in L^2(D)$.

Let a be a log-normal field as defined in Section 4.2.4 via a KL expansion and let a_K be the expansion truncated after K terms. Now, for almost all $\omega \in \Omega$, let $u_{K,h}(\omega,\cdot)$ denote the unique function in V_h that satisfies

$$b_{K,\omega}(u_{K,h}(\omega,\cdot), v_h) = L_\omega(v_h), \qquad \text{for all} \quad v_h \in V_h\,, \tag{4.19}$$

where $b_{K,\omega}(u,v) := \int_D a_K(\omega,x)\nabla u \cdot \nabla v \,\mathrm{d}x$. Then we have the following result.

Theorem 4.3.15. *Let $f \in L^{p^*}(\Omega, L^2(D))$, for some $p^* \in (0,\infty]$. Then, in the Gaussian covariance case, there exists a $c_1 > 0$ such that, for any $p < p^*$,*

$$\|u - u_{K,h}\|_{L^p(\Omega,H^1_0(D))} \lesssim \left(h \;+\; K^{\frac{d-1}{2d}} \exp\big(-c_1\, K^{1/d}\big)\right) \|f\|_{L^{p^*}(\Omega,L^2(D))}\,.$$

In the case of the 1-norm exponential covariance, for any $p < p^*$, $0 < \rho < 1$ and $0 < s < 1/2$,

$$\|u - u_{K,h}\|_{L^p(\Omega,H^1_0(D))} \lesssim \left(C_{4.3.15}\, h^s \;+\; K^{\frac{\rho-1}{2}}\right) \|f\|_{L^{p^*}(\Omega,L^2(D))}\,,$$

where $C_{4.3.15} := \min\{\eta(K)\, h^{1-s}, C_{4.3.9}\}$ and $\eta(K)$ is an exponential function of K given in [10, Proposition 6.3].

Proof. It follows from Proposition 4.2.6 that a_K satisfies assumptions A1–A2. Therefore we can apply Theorem 4.3.9 to bound $\|u_K - u_{K,h}\|_{L^2(\Omega,H^1_0(D))}$. The result then follows from Theorem 4.2.7, Proposition 4.2.8 and the triangle inequality. □

In contrast to [10, Proposition 6.3] this bound is uniform in K. Note that for small values of K, the constant $C_{4.3.15}$ will actually be of order $\mathcal{O}(h^{1-s})$ leading to a linear dependence on h also in the exponential covariance case (as stated in [10, Proposition 6.3]). For larger values of K this will not be the case and the lower regularity of u_K will affect the convergence rate with respect to h.

Remark 4.3.16. As we will see in Section 4.5, in the exponential covariance case for correlation lengths that are smaller than the diameter of the domain, a relatively large number K of KL-modes is necessary to get even 10% accuracy. The new uniform error bound in Theorem 4.3.15 is therefore crucial also for the analysis of stochastic Galerkin and stochastic collocation methods, such as the one given in [10].

4.4 Convergence Analysis of Multilevel Monte Carlo Methods

We will now apply this new finite element error analysis in Section 4.4, to give a rigorous bound on the cost of the multilevel Monte Carlo method applied to (4.1) for general random fields satisfying Assumptions A1–A3, and to establish its superiority over the classical Monte Carlo method. This builds on the recent paper [12]. We start by briefly recalling the classical Monte Carlo (MC) and multilevel Monte Carlo (MLMC) algorithms for PDEs with random coefficients, together with the main results on their performance. For a more detailed description of the methods, we refer the reader to [12] and the references therein.

In the Monte Carlo framework, we are usually interested in finding the expected value of some functional $Q = \mathcal{G}(u)$ of the solution u to our model problem (4.1). Since u is not easily accessible, Q is often approximated by the quantity $Q_h := \mathcal{G}(u_h)$, where u_h denotes the finite element solution on a sufficiently fine spatial grid $\mathcal{T}_h$. To simplify, in this section we will not consider quadrature error. Thus, to estimate $\mathbb{E}[Q]$, we compute approximations (or *estimators*) $\widehat{Q}_h$ to $\mathbb{E}[Q_h]$, and quantify the accuracy of our approximations via the root mean square error (RMSE)

$$e(\widehat{Q}_h) := \Big(\mathbb{E}\big[(\widehat{Q}_h - \mathbb{E}(Q))^2\big]\Big)^{1/2}.$$

The computational cost $\mathcal{C}_\varepsilon(\widehat{Q}_h)$ of our estimator is then quantified by the number of floating point operations that are needed to achieve a RMSE of $e(\widehat{Q}_h) \leq \varepsilon$. This will be referred to as the ε–cost.

The classical Monte Carlo (MC) estimator for $\mathbb{E}[Q_h]$ is

$$\widehat{Q}^{\mathrm{MC}}_{h,N} := \frac{1}{N}\sum_{i=1}^{N} Q_h(\omega^{(i)}), \tag{4.20}$$

where $Q_h(\omega^{(i)})$ is the ith sample of Q_h and N independent samples are computed in total.

There are two sources of error in the estimator (4.20), the approximation of Q by Q_h, which is related to the spatial discretisation, and the sampling error due to replacing the expected value by a finite sample average. This becomes clear when expanding the mean square error (MSE) and using the fact that for Monte Carlo $\mathbb{E}[\widehat{Q}^{\mathrm{MC}}_{h,N}] = \mathbb{E}[Q_h]$ and $\mathbb{V}[\widehat{Q}^{\mathrm{MC}}_{h,N}] = N^{-1}\,\mathbb{V}[Q_h]$, where $\mathbb{V}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$ denotes the variance of the random variable $X : \Omega \to \mathbb{R}$. We get

$$e(\widehat{Q}^{\mathrm{MC}}_{h,N})^2 \;=\; N^{-1}\mathbb{V}[Q_h] + \big(\mathbb{E}[Q_h - Q]\big)^2. \tag{4.21}$$

A sufficient condition to achieve a RMSE of ε with this estimator is that both of these terms are less than $\varepsilon^2/2$. For the first term, this is achieved by choosing a large enough number of samples, $N = \mathcal{O}(\varepsilon^{-2})$. For the second term, we need to choose a fine enough finite element mesh $\mathcal{T}_h$, such that $\mathbb{E}[Q_h - Q] = \mathcal{O}(\varepsilon)$.

The main idea of the MLMC estimator is very simple. We sample not just from one approximation Q_h of Q, but from several. Linearity of the expectation operator implies that

$$\mathbb{E}[Q_h] = \mathbb{E}[Q_{h_0}] + \sum_{\ell=1}^{L} \mathbb{E}[Q_{h_\ell} - Q_{h_{\ell-1}}] \tag{4.22}$$

where $\{h_\ell\}_{\ell=0,\dots,L}$ are the mesh widths of a sequence of increasingly fine triangulations $\mathcal{T}_{h_\ell}$ with $\mathcal{T}_h := \mathcal{T}_{h_L}$, the finest mesh, and $h_{\ell-1}/h_\ell \leq M^*$, for all $\ell = 1, \dots, L$. Hence, the expectation on the finest mesh is equal to the expectation on the coarsest mesh, plus a sum of corrections adding the difference in expectation between simulations on consecutive meshes. The multilevel idea is now to independently estimate each of these terms such that the overall variance is minimised for a fixed computational cost.

Setting for convenience $Y_0 := Q_{h_0}$ and $Y_\ell := Q_{h_\ell} - Q_{h_{\ell-1}}$, for $1 \leq \ell \leq L$, we define the MLMC estimator simply as

$$\widehat{Q}^{\mathrm{ML}}_{h,\{N_\ell\}} \;:=\; \sum_{\ell=0}^{L} \widehat{Y}^{\mathrm{MC}}_{\ell,N_\ell} \;=\; \sum_{\ell=0}^{L} \frac{1}{N_\ell}\sum_{i=1}^{N_\ell} Y_\ell(\omega^{(i)}), \tag{4.23}$$

where importantly $Y_\ell(\omega^{(i)}) = Q_{h_\ell}(\omega^{(i)}) - Q_{h_{\ell-1}}(\omega^{(i)})$, i.e. using the same sample on both meshes.

Since all the expectations $\mathbb{E}[Y_\ell]$ are estimated independently in (4.22), the variance of the MLMC estimator is $\sum_{\ell=0}^{L} N_\ell^{-1}\,\mathbb{V}[Y_\ell]$ and expanding as in (4.21) leads again to

$$e(\widehat{Q}^{\mathrm{ML}}_{h,\{N_\ell\}})^2 \;:=\; \mathbb{E}\Big[\big(\widehat{Q}^{\mathrm{ML}}_{h,\{N_\ell\}} - \mathbb{E}[Q]\big)^2\Big] \;=\; \sum_{\ell=0}^{L} N_\ell^{-1}\,\mathbb{V}[Y_\ell] \;+\; \big(\mathbb{E}[Q_h - Q]\big)^2. \tag{4.24}$$

As in the classical MC case before, we see that the MSE consists of two terms, the variance of the estimator and the error in mean between Q and Q_h. Note that the second term is identical to the second term for the classical MC method in (4.21). A sufficient condition to achieve a RMSE of ε is again to make both terms less than $\varepsilon^2/2$. This is easier to achieve with the MLMC estimator, as

– for sufficiently large h_0, samples of Q_{h_0} are much cheaper to obtain than samples of Q_h;
– the variance Y_ℓ tends to 0 as $h_\ell \to 0$, meaning we need fewer samples on $\mathcal{T}_{h_\ell}$, for $\ell > 0$.

Let now $\mathcal{C}_\ell$ denote the cost to obtain one sample of Q_{h_ℓ}. Then we have the following results on the ε–cost of the MLMC estimator (cf. [12, 35]).

Theorem 4.4.1. *Suppose that there are positive constants $\alpha, \beta, \gamma > 0$ such that $\alpha \geq \frac{1}{2}\min(\beta,\gamma)$ and*
M1. $|\mathbb{E}[Q_h - Q]| \;=\; O(h^\alpha)$
M2. $\mathbb{V}[Q_{h_\ell} - Q_{h_{\ell-1}}] \;=\; O(h_\ell^\beta)$
M2. $\mathcal{C}_\ell \;=\; O(h_\ell^{-\gamma})$,
Then, for any $\varepsilon < e^{-1}$, there exist a value L and a sequence $\{N_\ell\}_{\ell=0}^{L}$, such that $e(\widehat{Q}^{\mathrm{ML}}_{h,\{N_\ell\}}) < \varepsilon$ and

$$\mathcal{C}_\varepsilon(\widehat{Q}^{\mathrm{ML}}_{h,\{N_\ell\}}) \;\lesssim\; \begin{cases} \varepsilon^{-2}, & \text{if } \beta > \gamma, \\ \varepsilon^{-2}(\log\varepsilon)^2, & \text{if } \beta = \gamma, \\ \varepsilon^{-2-(\gamma-\beta)/\alpha}, & \text{if } \beta < \gamma. \end{cases}$$

For the classical MC estimator we have $\mathcal{C}_\varepsilon(\widehat{Q}^{\mathrm{MC}}_h) \;=\; O(\varepsilon^{-2-\gamma/\alpha})$.

We can now use the results from Section 4.3.2 to verify Assumptions M1–M2 in Theorem 4.4.1 for some simple functionals $\mathcal{G}(u)$. More complicated functionals could then be tackled by duality arguments in a similar way.

Proposition 4.4.2. *Let $Q = \mathcal{G}(u) := |u|^q_{H^1(D)}$, for some $1 \leq q < p_*/2$. Suppose that D is a $\mathcal{C}^2$ bounded domain and that Assumptions A1–A4 hold, for some $0 < t < 1$. Then Assumptions M1–M2 in Theorem 4.4.1 hold for any $\alpha < t$ and $\beta < 2t$. For $t = 1$, we get $\alpha = 1$ and $\beta = 2$.*

Proof. We only give the proof for $t < 1$. The proof for $t = 1$ is analogous. Let $Q_h := |u_h|^q_{H^1(D)}$. Using the expansion $a^q - b^q = (a-b)\sum_{j=0}^{q-1} a^j b^{q-1-j}$, for $a, b \in \mathbb{R}$ and $q \in \mathbb{N}$, we get

$$|Q(\omega) - Q_h(\omega)| \;\lesssim\; |u(\omega,\cdot) - u_h(\omega,\cdot)|_{H^1(D)} \max\Big\{|u(\omega,\cdot)|^{q-1}_{H^1(D)}, |u_h(\omega,\cdot)|^{q-1}_{H^1(D)}, 1\Big\}, \quad \text{almost surely.}$$

This also holds for non-integer values of $q > 1$. Now, it follows from Lemma 4.2.1 and Theorem 4.3.9 that, for any $\alpha < t$,

$$|Q(\omega) - Q_h(\omega)| \;\lesssim\; \frac{C_{4.3.1}(\omega)\, C_{4.3.8}(\omega)}{a^{q-1}_{\min}(\omega)}\, \|f(\omega,\cdot)\|^q_{H^{\alpha-1}(D)}\, h^\alpha, \quad \text{almost surely.}$$

Taking expectation on both sides and applying Hölder's inequality, since $q < p_*$, it follows from Assumptions A1–A3 that Assumption M1 holds with $\alpha < t$.

To prove Assumption M2, let us consider $Y_{h_\ell} := Q_{h_\ell} - Q_{h_{\ell-1}}$. As above, it follows from Lemma 4.2.1 and Theorem 4.3.9 together with the triangle inequality that, for any $s < t$,

$$\begin{aligned} |Y_{h_\ell}(\omega)| \;&\lesssim\; |u_{h_\ell}(\omega,\cdot) - u_{h_{\ell-1}}(\omega,\cdot)|_{H^1(D)} \max\Big\{|u_{h_\ell}(\omega,\cdot)|^{q-1}_{H^1(D)}, |u_{h_{\ell-1}}(\omega,\cdot)|^{q-1}_{H^1(D)}, 1\Big\} \\ &\lesssim\; \frac{C_{4.3.1}(\omega)\, C_{4.3.8}(\omega)}{a^{q-1}_{\min}(\omega)}\, \|f(\omega,\cdot)\|^q_{H^{s-1}(D)}\, h_\ell^s, \qquad \text{almost surely,} \end{aligned}$$

	$\lvert u\rvert_{H^1(D)}$		$\lVert u\rVert_{L^2(D)}$	
d	MC	MLMC	MC	MLMC
1	ε^{-3}	ε^{-2}	$\varepsilon^{-5/2}$	ε^{-2}
2	ε^{-4}	ε^{-2}	ε^{-3}	ε^{-2}
3	ε^{-5}	ε^{-3}	$\varepsilon^{-7/2}$	ε^{-2}

	$\lvert u\rvert_{H^1(D)}$		$\lVert u\rVert_{L^2(D)}$	
d	MC	MLMC	MC	MLMC
1	ε^{-4}	ε^{-2}	ε^{-3}	ε^{-2}
2	ε^{-6}	ε^{-4}	ε^{-4}	ε^{-2}
3	ε^{-8}	ε^{-6}	ε^{-5}	ε^{-3}

TAB. 4.1 – Theoretical upper bounds for the ε-costs of classical and multilevel Monte Carlo from Theorem 4.4.1 in the case of log-normal fields a with Gaussian covariance function (left) and exponential covariance function (right). (For simplicity we wrote ε^{-p}, instead of $\varepsilon^{-p-\delta}$ with $\delta > 0$.)

where the hidden constant depends on M^*. It follows again from Assumptions A1-A3 and Hölder's inequality, since $q < p_*/2$, that

$$\mathbb{V}[Y_{h_\ell}] \;=\; \mathbb{E}[Y_{h_\ell}^2] - (\mathbb{E}[Y_{h_\ell}])^2 \;\leq\; \mathbb{E}[Y_{h_\ell}^2] \;\lesssim\; h_\ell^\beta, \quad \text{where} \quad \beta < 2t.$$

□

Proposition 4.4.3. *Let $Q := \|u\|_{L^2(D)}^q$, for some $1 \leq q < p_*/2$. Suppose that D is a $\mathcal{C}^2$ bounded domain and that Assumptions A1–A4 hold, for some $0 < t < 1$. Then Assumptions M1–M2 in Theorem 4.4.1 hold for any $\alpha < 2t$ and $\beta < 4t$. For $t = 1$, we get $\alpha = 2$ and $\beta = 4$.*

Proof. This can be shown in the same way as Proposition 4.4.2 using Corollary 4.3.10 instead of Theorem 4.3.9. □

Substituting these values into Theorem 4.4.1 we can get theoretical upper bounds for the ε-costs of classical and multilevel Monte Carlo in the case of log-normal fields a, as shown in Table 4.1. We assume here that we can obtain individual samples in optimal cost $\mathcal{C}_\ell \lesssim h_\ell^{-d} \log(h_\ell^{-1})$ via a multigrid solver, i.e. $\gamma = d + \delta$ for any $\delta > 0$. We clearly see the advantages of the multilevel Monte Carlo method. More importantly, note that in the exponential covariance case in dimensions $d > 1$, the cost of MLMC is proportional to the cost of obtaining one sample on the finest grid, i.e. solving one deterministic PDE with the same regularity properties to accuracy ε. This implies that the method is optimal.

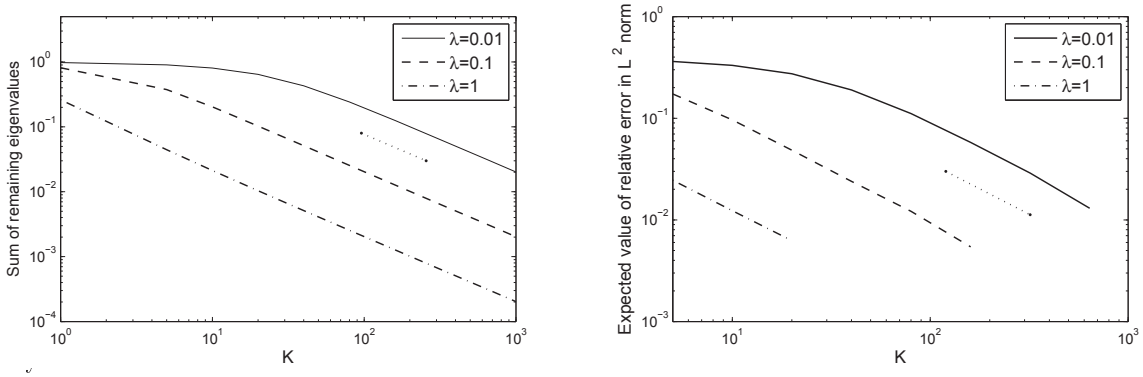
4.5 Numerical Results

In this section, we want to confirm numerically some of the results proved in earlier sections. All numerical results shown are for the case of a log-normal random coefficient $a(\omega, x)$ with exponential covariance function (4.6). All results are calculated for a model problem in 1D. The domain D is taken to be $(0, 1)$, and $f \equiv 1$. The sampling from the random coefficient $a(\omega, x)$ is done using a truncated KL-expansion, and as in the analysis in Section 4.3.3, we use the midpoint rule to approximate the integrals in the stiffness matrix. As the quantities of interest we will just study the simple functionals $Q := \|u\|_{L^2(D)}$ and $Q := |u|_{H^1(D)}$.

4.5.1 Convergence with Respect to K

We start with the results in Section 4.2.4. We want to show that in the exponential covariance case a relatively large number of KL-modes are necessary (even in 1D) to obtain acceptable accuracies especially when the correlation length is smaller than the diameter of the domain. In Figure 4.1 we study the decay of the eigenvalues corresponding to the KL-expansion, as well as the convergence with respect to K of $\mathbb{E}\left[\|u_K\|_{L^2(D)}\right]$ to $\mathbb{E}\left[\|u\|_{L^2(D)}\right]$. Since we do not know the exact solution and cannot solve the differential equation explicitly, we approximate u by u_{K^*,h^*} with $K^* = 5000$ and $h^* = 1/2048$ and u_K by u_{K,h^*} with $h^* = 1/2048$.

In the left figure we plot $\sum_{n>K} \theta_n$, i.e. the sum of the remaining eigenvalues when truncating after K terms. We see that after a short pre-asymptotic phase (depending on the size of λ) this sum decays linearly in K^{-1}. On the right we plot $\left|\mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)} - \|u_{K,h^*}\|_{L^2(D)}\right]\right| / \mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)}\right]$, the relative error. We see that the error decays roughly like $\sum_{n>K} \theta_n$ and thus also linearly in K^{-1}. This is better than



du

FIG. 4.1 – Let $d = 1$ and $\sigma^2 = 1$ in the exponential covariance case for different choices of the correlation length λ. Left : Plot of $\sum_{n>K} \theta_n$ as a function of K, i.e. the sum of the remaining KL eigenvalues when truncating after K terms. Right : The corresponding relative error $\left|\mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)} - \|u_{K,h^*}\|_{L^2(D)}\right]\right| / \mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)}\right]$ for problem (4.8) with $K^* = 5000$ and $h^* = 1/2048$. The dotted line has a gradient of -1.

the bound for $\mathbb{E}\left[\|u_{K^*,h^*} - u_{K,h^*}\|_{L^2(D)}\right] \geq \left|\mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)} - \|u_{K,h^*}\|_{L^2(D)}\right]\right|$ in Proposition 4.2.8, which predicts at most $\mathcal{O}(K^{-1/2})$ convergence, and it is related to weak convergence. As shown in [10, Section 5], the weak convergence order of the PDE solution with truncated KL-expansion can be $\mathcal{O}(\sum_{n>K} \theta_n)$ for certain functionals. We believe that the faster convergence of $\mathbb{E}\left[\|u_K\|_{L^2(D)}\right]$ to $\mathbb{E}\left[\|u\|_{L^2(D)}\right]$ can be shown using similar techniques.

4.5.2 Convergence with Respect to h

We now move on to the results in the main Section 4.3. For the reference solution we choose again $h^* = 1/2048$ and $K^* = 5000$. Figure 4.2 shows the convergence of $\mathbb{E}\left[|u_{K^*,h^*} - u_{K^*,h}|_{H^1(D)}\right]$ and $\mathbb{E}\left[\|u_{K^*,h^*} - u_{K^*,h}\|_{L^2(D)}\right]$ for (4.8). In the plots, a dash-dotted line indicates $\mathcal{O}(h^{1/2})$ convergence and a dotted line indicates $\mathcal{O}(h)$ convergence.

In both cases, we can see a short pre-asymptotic phase, which is again related to the correlation length λ. We can see in the right plot that the H^1–error converges with $\mathcal{O}(h^{1/2})$, confirming the sharpness of the bound proven in Theorem 4.3.13. The L^2–error converges linearly in h, and so the quadrature error does not seem to be dominant here.

4.5.3 Multilevel Monte Carlo Convergence

We first want to confirm Assumptions M1 and M2 in Theorem 4.4.1, for $Q = \|u\|_{L^2(D)}$. In other words, we want to confirm the rate of decay for the quantities $|\mathbb{E}\left[\|u_{K^*,h^*}\|_{L^2(D)} - \|u_{K^*,h}\|_{L^2(D)}\right]|$ and $\mathbb{V}\left[\|u_{K^*,h}\|_{L^2(D)} - \|u_{K^*,2h}\|_{L^2(D)}\right]$. Again, we choose $K^* = 5000$ and $h^* = 1/2048$ and look at the model problem (4.8) for $d = 1$. The results are shown in Figure 4.3. A dotted line indicates linear convergence, and a dashed line indicates quadratic convergence.

For the variance $\mathbb{V}\left[\|u_{K^*,h}\|_{L^2(D)} - \|u_{K^*,2h}\|_{L^2(D)}\right]$ (right plot), we observe the quadratic convergence predicted by Proposition 4.4.3. The expected value in the left plot seems to converge slightly faster than the predicted linear convergence. In our experience, this faster convergence depends on the choice of model problem and quantity of interest and will need to be investigated further, since it directly affects the cost of the multilevel Monte Carlo method through the size of the ratio β/α in the bound in Theorem 4.4.1.

The actual performance of the standard MC and MLMC estimators in estimating $Q = \|u\|_{L^2(D)}$ is shown in Figure 4.4. In the right plot the accuracy is scaled by the expected value of the quantity of interest. We see a clear advantage of the multilevel Monte Carlo method. For more numerical results, in particular results in 2D, we refer the reader to [12].

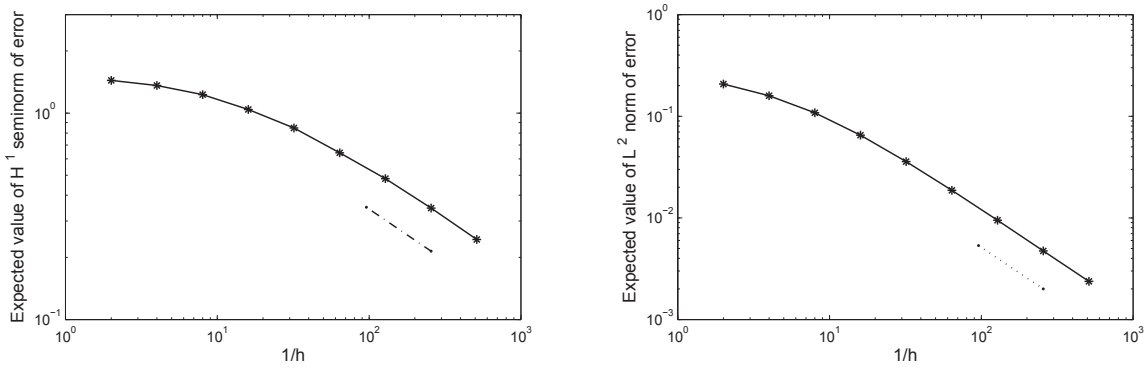


FIG. 4.2 – Left : Plot of $\mathbb{E}\left[|u_{K^*,h^*} - u_{K^*,h}|_{H^1(D)}\right]$ versus $1/h$ for model problem (4.8) with $d = 1$, $\lambda = 0.1$ and $\sigma^2 = 3$. Right : Corresponding L^2-error $\mathbb{E}\left[\|u_{K^*,h^*} - u_{K^*,h}\|_{L^2(D)}\right]$. The gradient of the dash-dotted (resp. dotted) line is $-1/2$ (resp. -1).

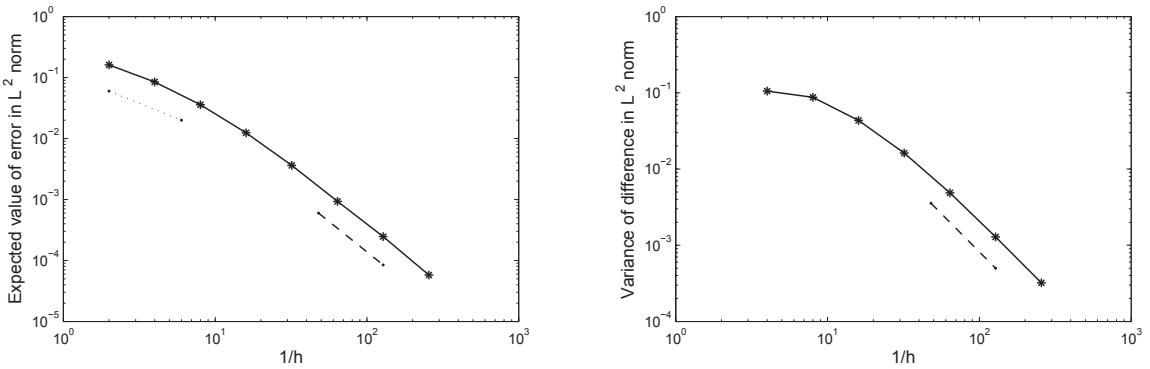


FIG. 4.3 – Left : Plot of $\left|\mathbb{E}\left[\|u_{K^*,h}\|_{L^2(D)} - \|u_{K^*,h^*}\|_{L^2(D)}\right]\right|$, for model problem (4.8) with $d = 1$, $\lambda = 0.1$, $\sigma^2 = 3$, $K^* = 5000$ and $h^* = 1/2048$. Right : Corresponding plot of the variance $\mathbb{V}\left[\|u_{K^*,h}\|_{L^2(D)} - \|u_{K^*,2h}\|_{L^2(D)}\right]$. The gradient of the dotted (resp. dashed) line is -1 (resp. -2).

4.6 Conclusions and Further Work

One of the major bottlenecks in probabilistic uncertainty quantification is the efficient numerical solution of PDEs with random coefficients. A typical model problem that arises in subsurface flow is an elliptic PDE with log-normal coefficients. The resulting PDE is not uniformly elliptic or bounded with respect to the random variable ω and usually lacks also full regularity. In this paper, we carry out for the first time a careful finite element error analysis under these weaker assumptions and provide optimal error estimates — with respect to the given regularity of the coefficients — for all finite moments of the finite element error (including quadrature errors). This then allows us to rigorously analyse the convergence and the cost of multilevel Monte Carlo methods, recently proposed for elliptic PDEs with random coefficients, in particular in the practically important case of log-normal coefficients with exponential covariance which has hitherto been unproved.

Via duality arguments, the results in this paper are readily applicable also to other functionals of the solution of more practical relevance. Other important aspects which we want to investigate in the future are the influence of the correlation length λ on the MLMC convergence, as well as the superconvergence in the expected value of certain model problems and quantities of interest (as highlighted in the discussion of Figure 4.2). Another issue which will require further research is the choice of quadrature rule. So far we

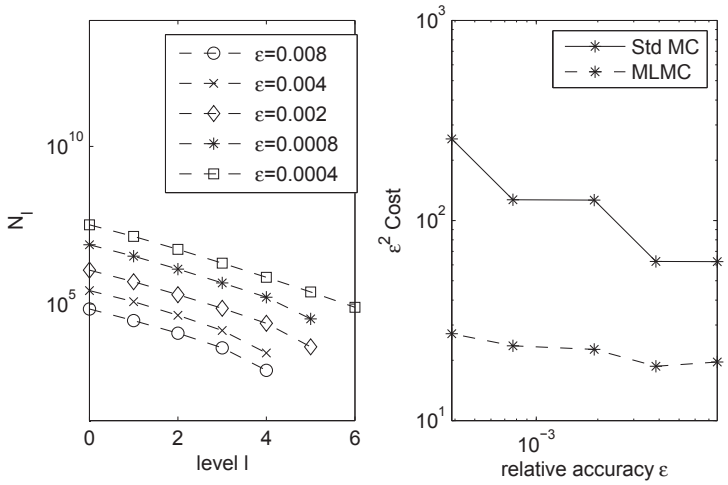


FIG. 4.4 – Left : Number of samples N_ℓ per level. Right : Plot of the cost scaled by ε^{-2} of the MLMC and standard MC estimators for $d = 1$, with $\lambda = 0.1$ and $\sigma^2 = 3$. The coarsest mesh size in all tests is $h_0 = 1/16$.

have only used simple midpoint or trapezoidal rules. In future work we would like to also explore multiscale methods and model reduction on coarser grids.

4.7 Appendix

In this appendix, we give more detailed proofs of Proposition 4.3.1 and Lemmas 4.3.2 and 4.3.3. The proofs follow those of Hackbusch [39, Theorems 9.1.8, 9.1.11 and 9.1.16], making explicit the dependence of all the constants that appear on the PDE coefficient a. The proof follows the classical Nirenberg translation method and consists in three main steps.

4.7.1 Step 1 – The Case $D = \mathbb{R}^d$

Proof of Lemma 4.3.2. In this proof we will use the norm on $H^s(\mathbb{R}^d)$ provided by the Fourier transform, $\|u\|^2_{H^s(\mathbb{R}^d)} := \|\hat{u}(\xi)(1+|\xi|^2)^{s/2}\|_{L^2(\mathbb{R}^d)}$, which is equivalent to the norm defined previously and defines the same space.

For any $h > 0$, we define the fractional difference operator (in direction $i = 1, \dots, d$) by

$$R^i_h(v)(x) := h^{-s} \sum_{\mu=0}^{+\infty} e^{-\mu h} (-1)^\mu \binom{s}{\mu} v(x + \mu h e_i),$$

where e_i is the ith unit vector in $\mathbb{R}^d$,

$$\binom{s}{0} = 1 \quad \text{and} \quad \binom{s}{\mu} (-1)^\mu = \frac{-s(1-s)(2-s)...(\mu-1-s)}{\mu!}.$$

Let us recall here some properties of R^i_h from [39, Proof of Theorem 9.1.8] :

- $(R^i_h)^*(v)(x) = h^{-s} \sum_{\mu=0}^{+\infty} e^{-\mu h} (-1)^\mu \binom{s}{\mu} v(x - \mu h e_i)$.
- For any $\tau \in \mathbb{R}$ and $v \in H^{\tau+s}(\mathbb{R}^d)$,

$$\|R^i_h v\|_{H^\tau(\mathbb{R}^d)} \leq \|v\|_{H^{\tau+s}(\mathbb{R}^d)} \text{ and } \|(R^i_h)^* v\|_{H^\tau(\mathbb{R}^d)} \leq \|v\|_{H^{\tau+s}(\mathbb{R}^d)}. \tag{4.25}$$

– $\widehat{R_h^i v}(\xi) = [(1-e^{-h+i\xi_j h})/h]^s \hat{v}(\xi)$ and $\widehat{(R_h^i)^* v}(\xi) = [(1-e^{-h-i\xi_j h})/h]^s \hat{v}(\xi)$.

We define for $u, v \in H^1(\mathbb{R}^d)$ the bilinear form

$$\begin{aligned} d(u,v) &:= \int_{\mathbb{R}^d} A\nabla u \nabla R_h^i(v)\,\mathrm{d}x - \int_{\mathbb{R}^d} A\nabla (R_h^i)^* u \nabla v\,\mathrm{d}x \\ &= \sum_{\mu=1}^{\infty} h^{-s} e^{-\mu h}(-1)^\mu \binom{s}{\mu} \int_{\mathbb{R}^d} (A(x-\mu h e_i) - A(x))\nabla u(x-\mu h e_i)\nabla v\,\mathrm{d}x. \end{aligned}$$

Hence,

$$|d(u,v)| \lesssim |A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))}\, |u|_{H^1(\mathbb{R}^d)}\, |v|_{H^1(\mathbb{R}^d)},$$

where the hidden constant is proportional to

$$\sum_{\mu=1}^{\infty} h^{-s} e^{-\mu h}(-1)^\mu \binom{s}{\mu} (\mu h)^t,$$

which is finite, since $\binom{s}{\mu} = \mathcal{O}(\mu^{-s-1})$ and thus $\sum_{\mu=1}^{\infty} e^{-\mu h}\mu^{t-s-1} = \mathcal{O}(h^{s-t})$. The three spaces $H^{1-s}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d) \subset H^{s-1}(\mathbb{R}^d)$ form a Gelfand triple, so that we can deduce, using (4.25), that

$$\begin{aligned} A_{\min}|(R_h^i)^* w|^2_{H^1(\mathbb{R}^d)} &\le \int_{\mathbb{R}^d} A\nabla (R_h^i)^* w \nabla (R_h^i)^* w\,\mathrm{d}x \\ &= -d(w, (R_h^i)^* w) + \langle F, R_h^i (R_h^i)^* w\rangle_{H^{s-1}(\mathbb{R}^d), H^{1-s}(\mathbb{R}^d)} \\ &\le |d(w,(R_h^i)^* w)| + \|F\|_{H^{s-1}(\mathbb{R}^d)} \|R_h^i (R_h^i)^* w\|_{H^{1-s}(\mathbb{R}^d)} \\ &\lesssim |A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))}\, |w|_{H^1(\mathbb{R}^d)}\, |(R_h^i)^* w|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)}\, \|(R_h^i)^* w\|_{H^1(\mathbb{R}^d)}, \end{aligned}$$

therefore we get

$$\begin{aligned} A_{\min}\|(R_h^i)^* w\|^2_{H^1(\mathbb{R}^d)} &\lesssim |A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))}\, |w|_{H^1(\mathbb{R}^d)}\, |(R_h^i)^* w|_{H^1(\mathbb{R}^d)} + \\ &\quad \|F\|_{H^{s-1}(\mathbb{R}^d)}\, \|(R_h^i)^* w\|_{H^1(\mathbb{R}^d)} + A_{\min}\|(R_h^i)^* w\|^2_{L^2(\mathbb{R}^d)} \\ &\lesssim |A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))}\, |w|_{H^1(\mathbb{R}^d)}\, |(R_h^i)^* w|_{H^1(\mathbb{R}^d)} + \\ &\quad \|F\|_{H^{s-1}(\mathbb{R}^d)}\, \|(R_h^i)^* w\|_{H^1(\mathbb{R}^d)} + A_{\min}\|(R_h^i)^* w\|_{H^{-1}(\mathbb{R}^d)} \|(R_h^i)^* w\|_{H^1(\mathbb{R}^d)}, \end{aligned}$$

and finally, using (4.25) once more,

$$\|(R_h^i)^* w\|_{H^1(\mathbb{R}^d)} \lesssim \frac{1}{A_{\min}} \Big(|A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))}\, |w|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)} \Big) + \|w\|_{L^2(\mathbb{R}^d)}.$$

For any $1 \ge h > 0$, since $|1-e^{-h-i\xi_i h}|^2 \ge |\mathrm{Im}(1-e^{-h-i\xi_i h})|^2 = e^{-2h}\sin(\xi_i h)^2 \ge e^{-2}\sin(\xi_i h)^2$, and since $\sin^2(\xi h) \ge \left(\frac{2}{\pi}\xi h\right)^2$, for all $|\xi| \le 1/h$, we have conversely that

$$\begin{aligned} \sum_{i=1}^{d} \|(R_h^i)^* w\|^2_{H^1(\mathbb{R}^d)} &\ge \int_{|\xi|\le 1/h} (1+|\xi|^2) \sum_{i=1}^{d} |\widehat{(R_h^i)^* w}(\xi)|^2 d\xi \\ &= \int_{|\xi|\le 1/h} (1+|\xi|^2) \sum_{i=1}^{d} \left| \frac{1-e^{-h-i\xi_i h}}{h} \right|^{2s} |\hat{w}(\xi)|^2 d\xi \\ &\ge e^{-2} \int_{|\xi|\le 1/h} (1+|\xi|^2) \sum_{i=1}^{d} \left| \frac{\sin(\xi_i h)}{h} \right|^{2s} |\hat{w}(\xi)|^2 d\xi \\ &\gtrsim \int_{|\xi|\le 1/h} (1+|\xi|^2) |\xi|^{2s} |\hat{w}(\xi)|^2 d\xi. \end{aligned}$$

Hence, for any $0 < h \leq 1$, we obtain

$$\begin{aligned}\|w\|^2_{H^{1+s}(\mathbb{R}^d)} &\leq \int_{\mathbb{R}^d}(1+|\xi|^2)|\xi|^{2s}|\hat{w}(\xi)|^2 d\xi \;+\; \int_{\mathbb{R}^d}(1+|\xi|^2)|\hat{w}(\xi)|^2 d\xi \\ &\leq \sum_{i=1}^{d}\|(R_h^i)^* w\|^2_{H^1(\mathbb{R}^d)} \;+\; \|w\|^2_{H^1(\mathbb{R}^d)}\end{aligned}$$

and so

$$\|w\|_{H^{1+s}(\mathbb{R}^d)} \lesssim \frac{1}{A_{\min}}\Big(|A|_{\mathcal{C}^t(\mathbb{R}^d, S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d)} + \|F\|_{H^{s-1}(\mathbb{R}^d)}\Big) + \|w\|_{H^1(\mathbb{R}^d)} \;<\; +\infty.$$

□

4.7.2 Step 2 – The Case $D = \mathbb{R}^d_+$

In order to proof Lemma 4.3.3 we will need the following two lemmas.

Lemma 4.7.1 ([39, Lemma 9.1.12]). *Let $s > 0$, $s \neq 1/2$, then*

$$|||v|||_s := \left(\|v\|^2_{L^2(\mathbb{R}^d_+)} + \sum_{i=1}^{d}\left\|\frac{\partial v}{\partial x_i}\right\|^2_{H^{s-1}(\mathbb{R}^d_+)}\right)^{1/2}$$

defines a norm on $H^s(\mathbb{R}^d_+)$ that is equivalent to $\|v\|_{H^s(\mathbb{R}^d_+)}$.

Lemma 4.7.2. *Let $D \subset \mathbb{R}^d$ and $0 < s < t < 1$. If $b \in \mathcal{C}^t(\overline{D})$ and $v \in H^s(D)$, then $bv \in H^s(D)$ and*

$$\|bv\|_{H^s(D)} \lesssim |b|_{\mathcal{C}^t(\overline{D})}\|v\|_{L^2(D)} \;+\; |b|_{\mathcal{C}^0(\overline{D})}\|v\|_{H^s(D)}.$$

The hidden constant depends only on t, s and d.

Proof. This is a classical result, but we require again the exact dependence of the bound on b. First note that trivially $\|bv\|_{L^2(D)} \leq b_{\max}\|v\|_{L^2(D)}$ where $b_{\max} := |b|_{\mathcal{C}^0(\overline{D})}$. Now, for any $x, y \in D$, we have

$$|b(x)v(x) - b(y)v(y)|^2 \leq 2(b(x)^2|v(x)-v(y)|^2 + v(y)^2|b(x)-b(y)|^2).$$

Continuing v by 0 on $\mathbb{R}^d$ and denoting this extension by $\tilde{v}$, this implies

$$\begin{aligned}\iint_{D^2}\frac{|b(x)v(x)-b(y)v(y)|^2}{\|x-y\|^{d+2s}}\,\mathrm{d}x\,\mathrm{d}y &\leq 2b^2_{\max}\|v\|^2_{H^s(D)} + 2\iint_{D^2}\frac{v(y)^2|b(x)-b(y)|^2}{\|x-y\|^{d+2s}} \\ &\leq 2b^2_{\max}\|v\|^2_{H^s(D)} + \iint_{\substack{x,y\in D\\ \|x-y\|\geq 1}} 8b^2_{\max}\frac{v(y)^2}{\|x-y\|^{d+2s}} \;+\; 2|b|^2_{\mathcal{C}^t(\overline{D})}\frac{v(y)^2}{\|x-y\|^{d+2(s-t)}}\,\mathrm{d}x\,\mathrm{d}y \\ &\leq 2b^2_{\max}\|v\|^2_{H^s(D)} + \left(8b^2_{\max}\left\|\frac{\mathbb{1}_{\|z\|\geq 1}}{\|z\|^{d+2s}}\right\|_{L^1(\mathbb{R}^d)} + 2|b|^2_{\mathcal{C}^t(\overline{D})}\left\|\frac{\mathbb{1}_{\|z\|\leq 1}}{\|z\|^{d+2(s-t)}}\right\|_{L^1(\mathbb{R}^d)}\right)\|\tilde{v}\|^2_{L^2(\mathbb{R}^d)} \\ &\lesssim b^2_{\max}\|v\|^2_{H^s(D)} \;+\; |b|^2_{\mathcal{C}^t(\overline{D})}\|v\|^2_{L^2(D)}.\end{aligned}$$

□

Proof of Lemma 4.3.3. First we continue the solution w by 0 on $\mathbb{R}^d \setminus \mathbb{R}^d_+$ and denote the extension $\tilde{w} \in H^1(\mathbb{R}^d)$. Take $1 \leq i \leq d-1$. Similarly to the previous section, we define for $u, v \in H^1(\mathbb{R}^d)$

$$d(u,v) \;:=\; \int_{\mathbb{R}^d_+} A\nabla u \nabla R^i_h(v)\,\mathrm{d}x - \int_{\mathbb{R}^d_+} A\nabla (R^i_h)^* u \nabla v\,\mathrm{d}x$$

and deduce again that

$$|d(u,v)| \lesssim |A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}\,|u|_{H^1(\mathbb{R}^d_+)}\,|v|_{H^1(\mathbb{R}^d_+)}.$$

We now note that, since $i \neq d$, $(R_h^i)^* w \in H_0^1(\mathbb{R}^d_+)$ and $(R_h^i)^*\tilde{w} \in H^1(\mathbb{R}^d)$ is equal to the continuation by 0 on $\mathbb{R}^d \setminus \mathbb{R}^d_+$ of $(R_h^i)^* w$. We deduce, similarly to the proof in Section 4.7.1 using (4.25), that

$$\|(R_h^i)^*\tilde{w}\|_{H^1(\mathbb{R}^d)} \lesssim \frac{1}{A_{\min}}\Big(|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d_+)} + \|F\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) + \|w\|_{H^1(\mathbb{R}^d_+)} =: \mathcal{B}(w).$$

(Note that we added $|w|_{H^1(\mathbb{R}^d_+)}$ to the bound to simplify the notation later.) Hence, by the same token as in the previous section, we get

$$\int_{\mathbb{R}^d}(1+|\xi|^2)(|\xi_1|^2+\ldots+|\xi_{d-1}|^2)^s|\widehat{\tilde{w}}(\xi)|^2 d\xi \lesssim \mathcal{B}(w)^2. \tag{4.26}$$

In particular, this implies that, for $1 \le i \le d$ and $1 \le j \le d-1$, we have

$$\int_{\mathbb{R}^d}\left|\widehat{\frac{\partial^2\tilde{w}}{\partial x_i \partial x_j}}(\xi)\right|^2 (1+|\xi|^2)^{s-1} d\xi = \int_{\mathbb{R}^d}|\xi_i|^2|\xi_j|^2|\widehat{\tilde{w}}(\xi)|^2(1+|\xi|^2)^{s-1}d\xi \lesssim \mathcal{B}(w)^2,$$

which means that $\frac{\partial^2\tilde{w}}{\partial x_i\partial x_j} \in H^{s-1}(\mathbb{R}^d)$ and $\left\|\frac{\partial^2\tilde{w}}{\partial x_i\partial x_j}\right\|_{H^{s-1}(\mathbb{R}^d)} \lesssim \mathcal{B}(w)$. In particular, for all $(i,j) \neq (d,d)$, this further implies that $\frac{\partial^2 w}{\partial x_i\partial x_j} \in H^{s-1}(\mathbb{R}^d_+)$ and that $\left\|\frac{\partial^2 w}{\partial x_i\partial x_j}\right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim \mathcal{B}(w)$. Using Lemma 4.7.1 we deduce that $\frac{\partial w}{\partial x_j} \in H^s(\mathbb{R}^d_+)$ and that

$$\left\|\frac{\partial w}{\partial x_j}\right\|_{H^s(\mathbb{R}^d_+)} \lesssim \mathcal{B}(w), \qquad \text{for all } 1 \le j \le d-1. \tag{4.27}$$

It remains to bound $\left\|\frac{\partial w}{\partial x_d}\right\|_{H^s(\mathbb{R}^d_+)}$, which is rather technical. To achieve it we will use the PDE (4.9), Lemma 4.7.2 and the following result.

Lemma 4.7.3. *For almost all $x_d \in \mathbb{R}$, we have $\frac{\partial\tilde{w}}{\partial x_d}(.,x_d) \in H^s(\mathbb{R}^{d-1})$ and*

$$\int_{\mathbb{R}}\left\|\frac{\partial\tilde{w}}{\partial x_d}(.,x_d)\right\|^2_{H^s(\mathbb{R}^{d-1})} \mathrm{d}x_d = \int_{\mathbb{R}^d}(1+|\xi'|^2)^s|\xi_d|^2\left|\widehat{\tilde{w}}(\xi)\right|^2 d\xi \lesssim \mathcal{B}(w)^2.$$

Proof. This follows from Fubini's theorem and Plancherel's formula, together with (4.26). □

From this we deduce that $\frac{\partial w}{\partial x_d}(.,x_d) \in H^s(\mathbb{R}^{d-1})$, for almost all $x_d \in \mathbb{R}_+$, and that

$$\int_{\mathbb{R}_+}\left\|\frac{\partial w}{\partial x_d}(.,x_d)\right\|^2_{H^s(\mathbb{R}^{d-1})} \mathrm{d}x_d \lesssim \mathcal{B}(w)^2.$$

Let $1 \le i \le d-1$. Using Lemma 4.7.2 we deduce that $A_{id}\frac{\partial w}{\partial x_d}(.,x_d) \in H^s(\mathbb{R}^{d-1})$, for almost all $x_d \in \mathbb{R}_+$, and that

$$\begin{aligned}\left\|\left(A_{id}\frac{\partial w}{\partial x_d}\right)(.,x_d)\right\|_{H^s(\mathbb{R}^{d-1})} &\lesssim |A_{id}(.,x_d)|_{\mathcal{C}^t(\mathbb{R}^{d-1})}\left\|\frac{\partial w}{\partial x_d}(.,x_d)\right\|_{L^2(\mathbb{R}^{d-1})} \\ &+ \|A_{id}(.,x_d)\|_{\mathcal{C}^0(\mathbb{R}^{d-1})}\left\|\frac{\partial w}{\partial x_d}(.,x_d)\right\|_{H^s(\mathbb{R}^{d-1})}.\end{aligned}$$

Therefore, since by definition $\|A_{id}\|_{\mathcal{C}^0(\mathbb{R}^d)} \le A_{\max}$, we get

$$\int_{\mathbb{R}_+}\left\|A_{id}\frac{\partial w}{\partial x_d}\right\|^2_{H^s(\mathbb{R}^{d-1})} \mathrm{d}x_d \lesssim |A_{id}|^2_{\mathcal{C}^t(\overline{\mathbb{R}^d_+})}\,|w|^2_{H^1(\mathbb{R}^d_+)} + A^2_{\max}\,\mathcal{B}(w)^2 \lesssim A^2_{\max}\,\mathcal{B}(w)^2.$$

Since $\frac{\partial}{\partial x_i}$ is linear continuous from $H^{1-s}(\mathbb{R}^{d-1})$ to $H^{-s}(\mathbb{R}^{d-1})$ (cf. [39, Remark 6.3.14(b)]) we can deduce from this that $\frac{\partial}{\partial x_i}\left(A_{id}\frac{\partial w}{\partial x_d}\right) \in H^{s-1}(\mathbb{R}^d_+)$ and that

$$\left\|\frac{\partial}{\partial x_i}\left(A_{id}\frac{\partial w}{\partial x_d}\right)\right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\,\mathcal{B}(w), \qquad \text{for all } 1\le i\le d-1. \tag{4.28}$$

To see this take $\varphi\in\mathcal{D}(\mathbb{R}^d_+)$. Then

$$\begin{aligned}\left|\left\langle\frac{\partial}{\partial x_i}\left(A_{id}\frac{\partial w}{\partial x_d}\right),\varphi\right\rangle_{\mathcal{D}'(\mathbb{R}^d_+),\mathcal{D}(\mathbb{R}^d_+)}\right| &= \left\|A_{id}\frac{\partial w}{\partial x_d}(x',x_d)\,\frac{\partial\varphi}{\partial x_i}(x',x_d)\right\|_{L^2(\mathbb{R}_+,H^s(\mathbb{R}^{d-1}))}\\ &\le \left\|A_{id}\frac{\partial w}{\partial x_d}\right\|_{L^2(\mathbb{R}_+,H^s(\mathbb{R}^{d-1}))}\left\|\frac{\partial\varphi}{\partial x_i}(x',x_d)\right\|_{L^2(\mathbb{R}_+,H^{-s}(\mathbb{R}^{d-1}))}\\ &\le \left\|A_{id}\frac{\partial w}{\partial x_d}\right\|_{L^2(\mathbb{R}_+,H^s(\mathbb{R}^{d-1}))}\|\varphi\|_{L^2(\mathbb{R}_+,H^{1-s}(\mathbb{R}^{d-1}))}.\end{aligned}$$

Using (4.27) and Lemma 4.7.2, we deduce in a similar way that $\frac{\partial}{\partial x_i}\left(A_{ij}\frac{\partial w}{\partial x_j}\right)\in H^{s-1}(\mathbb{R}^d_+)$ and that

$$\left\|\frac{\partial}{\partial x_i}\left(A_{ij}\frac{\partial w}{\partial x_j}\right)\right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\,\mathcal{B}(w), \quad \text{for all } 1\le i\le d \text{ and } 1\le j\le d-1. \tag{4.29}$$

We can now use the PDE (4.9) to get a similar bound for $(i,j)=(d,d)$. Since $F\in H^{s-1}(\mathbb{R}^d_+)$, it follows from (4.28) and (4.29) that $\frac{\partial}{\partial x_d}\left(A_{dd}\frac{\partial w}{\partial x_d}\right)\in H^{s-1}(\mathbb{R}^d_+)$ and that

$$\left\|\frac{\partial}{\partial x_d}\left(A_{dd}\frac{\partial w}{\partial x_d}\right)\right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\,\mathcal{B}(w) \;+\; \|F\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\,\mathcal{B}(w)\,.$$

Analogously to (4.28) we can prove that

$$\left\|\frac{\partial}{\partial x_i}\left(A_{dd}\frac{\partial w}{\partial x_d}\right)\right\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim A_{\max}\,\mathcal{B}(w), \qquad \text{for all } 1\le i\le d-1.$$

Hence, we can finally apply Lemma 4.7.1 to get that $A_{dd}\frac{\partial w}{\partial x_d}\in H^s(\mathbb{R}^d_+)$ and that

$$\left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|^2_{H^s(\mathbb{R}^d_+)} \lesssim \sum_{i=1}^d\left\|\frac{\partial}{\partial x_i}\left(A_{dd}\frac{\partial w}{\partial x_d}\right)\right\|^2_{H^{s-1}(\mathbb{R}^d_+)} + \left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|^2_{L^2(\mathbb{R}^d_+)} \lesssim A^2_{\max}\,\mathcal{B}(w)^2\,.$$

By applying Lemma 4.7.2 again, this time with $b:=1/A_{dd}$ and $v:=A_{dd}\frac{\partial w}{\partial x_d}$, we deduce that $\frac{\partial w}{\partial x_d}\in H^s(\mathbb{R}^d_+)$ and that

$$\begin{aligned}\left\|\frac{\partial w}{\partial x_d}\right\|_{H^s(\mathbb{R}^d_+)} &\lesssim \left|\frac{1}{A_{dd}}\right|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+})}\left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|_{L^2(\mathbb{R}^d_+)} + \frac{1}{A_{\min}}\left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|_{H^s(\mathbb{R}^d_+)}\\ &\lesssim \frac{|A_{dd}|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+})}}{A^2_{\min}}A_{\max}\,|w|_{H^1(\mathbb{R}^d_+)} + \frac{1}{A_{\min}}\left\|A_{dd}\frac{\partial w}{\partial x_d}\right\|_{H^s(\mathbb{R}^d_+)} \lesssim \frac{A_{\max}}{A_{\min}}\,\mathcal{B}(w)\,.\end{aligned}$$

To finish the proof we use this bound together with (4.27) and apply once more Lemma 4.7.1 to show that $w\in H^{1+s}(\mathbb{R}^d_+)$ and

$$\|w\|_{H^{1+s}(\mathbb{R}^d_+)} \lesssim \frac{A_{\max}}{A^2_{\min}}\Big(|A|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}\,|w|_{H^1(\mathbb{R}^d_+)} + \|F\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) \;+\; \frac{A_{\max}}{A_{\min}}\|w\|_{H^1(\mathbb{R}^d_+)}.$$

□

4.7.3 Step 3 – The Case D Bounded

We can now prove Proposition 4.3.1 using Lemmas 4.3.2 and 4.3.3 in two successive steps. We recall that D was assumed to be $\mathcal{C}^2$. Let $(D_i)_{0\le i\le m}$ be a covering of D such that the $(D_i)_{0\le i\le m}$ are open and bounded, $\overline{D}\subset\cup_{i=0}^m D_i$, $\cup_{i=1}^m (D_i\cap\partial D)=\partial D$, $\overline{D}_0\subset D$.

Let $(\chi_i)_{0\le i\le m}$ be a partition of unity subordinate to this cover, i.e. we have $\chi_i\in\mathcal{C}^\infty(\mathbb{R}^d,\mathbb{R}_+)$ with compact support $\mathrm{supp}(\chi_i)\subset D_i$, such that $\sum_{i=0}^p\chi_i=1$ on $\overline{D}$. We denote by u the solution of (4.3) and split it into $u=\sum_{i=0}^p u_i$, with $u_i=u\chi_i$. We treat now separately u_0 and then u_i, $1\le i\le p$, using Section 4.7.1 and 4.7.2, respectively.

Lemma 4.7.4. *u_0 belongs to $H^{1+s}(D)$ and*

$$\|u_0\|_{H^{1+s}(D)} \lesssim \frac{\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}^2}\|f\|_{H^{s-1}(D)}.$$

Proof. Since $\mathrm{supp}(u_0)\subset D_0$, we have that $u_0\in H_0^1(D)$ and it is the weak solution of the new equation $-\mathrm{div}(a\nabla u_0)=F$ on D, where

$$F := f\chi_0+a\nabla u\cdot\nabla\chi_0+\mathrm{div}(au\nabla\chi_0) \quad \text{on } D.$$

To apply Lemma 4.3.2 we will now extend all terms to $\mathbb{R}^d$, but continue to denote them the same. The terms u_0 and $f\chi_0+a\nabla u\cdot\nabla\chi_0$ can both be extended by 0. Their extensions will belong to $H^1(\mathbb{R}^d)$ and $H^{s-1}(\mathbb{R}^d)$, respectively. It follows from Lemma 4.7.2 that $au\in H^s(D)$ and so if we continue $au\nabla\chi_0$ by 0 on $\mathbb{R}^d$, the extension belongs to $H^s(\mathbb{R}^d)$, since $\mathrm{supp}(\chi_0)$ is compact in D. Using the fact that div is linear and continuous from $H^s(\mathbb{R}^d)$ to $H^{s-1}(\mathbb{R}^d)$ (cf. [39, Remark 6.3.14(b)] and the proof of (4.28) above), we can deduce that the divergence of the extension of $au\nabla\chi_0$ is in $H^{s-1}(\mathbb{R}^d)$, leading to an extension of F on $\mathbb{R}^d$, which belongs to $H^{s-1}(\mathbb{R}^d)$.

Let $\psi\in\mathcal{C}^\infty(\mathbb{R}^d,[0,1])$ such that $\psi=0$ on D_0 and $\psi=1$ on $\tilde{D}^c$, where $\tilde{D}$ is an open set such that $\overline{D}_0\subset\tilde{D}$ and $\overline{\tilde{D}}\subset D$. We use the following extension of a from D_0 to all of $\mathbb{R}^d$:

$$\overline{a}(x) := \begin{cases} a(x)(1-\psi(x))+a_{\min}\psi(x), & \text{if } x\in D,\\ a_{\min}\psi(x), & \text{otherwise.}\end{cases}$$

This implies that $\overline{a}\in\mathcal{C}^t(\mathbb{R}^d)$, and for any $x\in\mathbb{R}^d$, $a_{\min}\le\overline{a}(x)\le a(x)$ and $|\overline{a}|_{\mathcal{C}^t(\mathbb{R}^d)}\lesssim\|a\|_{\mathcal{C}^t(\overline{D})}$.

Using these extensions, we have that $-\mathrm{div}(\overline{a}\nabla u_0)=F$ in $\mathcal{D}'(\mathbb{R}^d)$. Indeed, for any $v\in\mathcal{D}(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d}\overline{a}(x)\nabla u_0(x)\nabla v(x) = \int_D a(x)\nabla u_0(x)\nabla v(x) \qquad \text{for any } v\in\mathcal{D}(\mathbb{R}^d),$$

since $\mathrm{supp}(u_0)$ is included in the open bounded set D_0, which implies that $\nabla u_0=0$ on D_0^c and $a=\overline{a}$ on D_0. Since $u\in H_0^1(D)$, we have by Poincaré's inequality that $\|u\|_{L^2(D)}\lesssim|u|_{H^1(D)}$. Therefore it follows from Lemma 4.2.1 that

$$|u_0|_{H^1(\mathbb{R}^d)} \le |u|_{H^1(D)}\|\chi_0\|_\infty+\|u\|_{L^2(D)}\|\nabla\chi_0\|_\infty \lesssim \frac{\|f\|_{H^{s-1}(D)}}{a_{\min}}.$$

Since $\chi_0\in\mathcal{C}^\infty(\mathbb{R}^d)$, using Lemma 4.7.2 and the linearity and continuity of div from $H^s(\mathbb{R}^d)$ to $H^{1-s}(\mathbb{R}^d)$ we further get

$$\begin{aligned}\|F\|_{H^{s-1}(\mathbb{R}^d)} &\le \|f\chi_0\|_{H^{s-1}(\mathbb{R}^d)}+\|a\nabla u\cdot\nabla\chi_0\|_{H^{s-1}(\mathbb{R}^d)}+\|\mathrm{div}(au\nabla\chi_0)\|_{H^{s-1}(\mathbb{R}^d)}\\ &\lesssim \|f\|_{H^{s-1}(D)}+a_{\max}|u|_{H^1(D)}+|a|_{\mathcal{C}^t(D)}\|u\|_{L^2(D)}+a_{\max}\|u\|_{H^s(D)}\\ &\lesssim \frac{\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}}\|f\|_{H^{s-1}(D)}.\end{aligned}$$

We can now apply Lemma 4.3.2 with $A=\overline{a}I_d$ and $w=u_0$ to show that $u_0\in H^{1+s}(\mathbb{R}^d)$ and

$$\|u_0\|_{H^{1+s}(\mathbb{R}^d)} \lesssim \frac{1}{a_{\min}}\Big(|\overline{a}|_{\mathcal{C}^t(\overline{D})}|u_0|_{H^1(\mathbb{R}^d)}+\|F\|_{H^{s-1}(\mathbb{R}^d)}\Big)+\|u_0\|_{H^1(\mathbb{R}^d)} \lesssim \frac{\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}^2}\|f\|_{H^{s-1}(D)}.$$

The hidden constant depends on the choices of χ_0 and ψ and on the constant in Poincaré's inequality, which depends on the shape and size of D, but not on a. □

Let us now treat the case of u_i, $1 \le i \le m$.

Lemma 4.7.5. *For $1 \le i \le m$, $u_i \in H^{1+s}(D)$ and*

$$\|u_i\|_{H^{1+s}(D)} \lesssim \frac{a_{\max}\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}^3}\|f\|_{H^{s-1}(D)}.$$

Proof. Similarly to the proof of the previous Lemma, $u_i \in H_0^1(D \cap D_i)$ is the weak solution of a new problem $-\mathrm{div}(a\nabla u_i) = g_i$ in $\mathcal{D}'(D \cap D_i)$ with

$$g_i := f\chi_i + a\nabla u \cdot \nabla\chi_i + \mathrm{div}(au\nabla\chi_i).$$

As in Lemma 4.7.4, since again div is linear and continuous from H^s to H^{s-1}, we can establish that $g_i \in H^{s-1}(D \cap D_i)$ and

$$\|g_i\|_{H^{s-1}(D\cap D_i)} \lesssim \frac{\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}}\|f\|_{H^{s-1}(D)}.$$

Now let $Q = \{(y', y_d) \in \mathbb{R}^{d-1} \times \mathbb{R} : |y'| < 1 \text{ and } |y_d| < 1\}$, $Q_0 = \{(y', y_d) \in \mathbb{R}^{d-1} \times \{0\} |\ \|y'\| < 1\}$ and $Q_+ = Q \cap \mathbb{R}^d_+$. For $1 \le i \le p$, let α_i be a bijection from D_i to Q such that $\alpha_i \in \mathcal{C}^2(\overline{D_i})$, $\alpha_i^{-1} \in \mathcal{C}^2(\overline{Q})$, $\alpha_i(D_i \cap D) = Q_+$ and $\alpha_i(D_i \cap \partial D) = Q_0$.

For all $y \in Q_+$, we define $w_i(y) := u_i(\alpha_i^{-1}(y)) \in H_0^1(Q_+)$ with $\nabla w_i(y) = J_i^{-t}(y)\nabla u_i(\alpha_i^{-1}(y))$, where $J_i(y) := D\alpha_i(\alpha_i^{-1}(y)))$ is the Jacobian of α_i. Furthermore, for $x \in D_i \cap D$ and $\varphi \in H_0^1(Q_+)$, we define $v(x) := \varphi(\alpha_i(x))$. Then $v \in H_0^1(D_i \cap D)$ and $\nabla v(\alpha_i^{-1}(y)) = J_i^t(y)\nabla\varphi(y)$, for all $y \in Q_+$, so that

$$\begin{aligned}\int_{D_i\cap D} a(x)\nabla u_i(x)\cdot\nabla v(x)\,\mathrm{d}x &= \int_{Q_+} a(\alpha_i^{-1}(y))\,\nabla u_i(\alpha_i^{-1}(y))\cdot\nabla v(\alpha_i^{-1}(y))\,|\mathrm{det}J_i(y))|^{-1}\,\mathrm{d}y \\ &= \int_{Q_+} A_i(y)\,\nabla w_i(y)\cdot\nabla\varphi(y)\,\mathrm{d}y\,,\end{aligned}$$

where

$$A_i(y) := a(\alpha_i^{-1}(y))\,|\mathrm{det}J_i(y))|^{-1}(J_i\,J_i^t)(y) \in S_d(\mathbb{R}).$$

We define $F_i \in H^{s-1}(Q_+)$ by

$$\langle F_i, \varphi\rangle_{H^{s-1}(Q_+),H_0^{1-s}(Q_+)} := \langle g_i, \varphi\circ\alpha_i\rangle_{H^{s-1}(D_i\cap D),H_0^{1-s}(D_i\cap D)}, \qquad \text{for all } \varphi \in H_0^{1-s}.$$

Indeed, since we assumed that α_i and α_i^{-1} are in $\mathcal{C}^2$, we have $\varphi\circ\alpha_i \in H_0^{1-s}(D_i \cap D)$ and moreover $\|\varphi\circ\alpha_i\|_{H^{1-s}(D_i\cap D)} \lesssim \|\varphi\|_{H^{1-s}(Q_+)}$ (cf. [39, Theorems 6.2.17 and 6.2.25(g)]), which implies that $F_i \in H^{s-1}(Q_+)$ and

$$\|F_i\|_{H^{s-1}(Q_+)} \lesssim \|g_i\|_{H^{s-1}(D\cap D_i)}.$$

We finally get that $v_i \in H_0^1(Q_+)$ solves

$$\int_{Q_+} A_i\nabla v_i\cdot\nabla\varphi\,\mathrm{d}y = \langle F_i,\varphi\rangle_{H^{s-1}(Q_+),H_0^{1-s}(Q_+)} \qquad \text{for all } \varphi \in H_0^1(Q_+).$$

In order to apply Lemma 4.3.3 we check first that $A_i \in \mathcal{C}^t(\overline{Q_+}, S_d(\mathbb{R}))$ and that it is coercive, and then define an extension of A_i to $\mathbb{R}^d_+$. Recalling that α_i is a $\mathcal{C}^2$–diffeomorphism from $D_i \cap D$ to Q_+, with $\alpha_i^{-1} \in \mathcal{C}^2(\overline{Q_+})$, we have for any $y \in Q_+$ and $\xi \in \mathbb{R}^d$:

– Coercivity : $A_i(y)\xi\cdot\xi = a(\alpha_i^{-1}(y))|\mathrm{det}J_i(y)|^{-1}\,|J_i^t(y)\xi|^2 \gtrsim a_{\min}|\xi|^2$. Hence $A_{\min} \gtrsim a_{\min}$.

– Boundedness :

$$A_{\max} := \|A_i\|_{\mathcal{C}^0(\overline{Q_+},S_d(\mathbb{R}))} = \max_{x\in\overline{D_i\cap D}} a(x)\,\||\mathrm{det}J_i|^{-1}J_iJ_i^t\|_{\mathcal{C}^0(\overline{Q_+},S_d(\mathbb{R}))} \lesssim a_{\max}.$$

– Regularity : $A_i \in \mathcal{C}^t(\overline{Q_+}, S_d(\mathbb{R}))$ and

$$\begin{aligned}\|A_i\|_{\mathcal{C}^t(\overline{Q_+},S_d(\mathbb{R}))} &\leq a_{\max}\,\||\det J_i|^{-1}J_iJ_i^t\|_{\mathcal{C}^t(\overline{Q_+},S_d(\mathbb{R}))} \\ &+ |a|_{\mathcal{C}^t(\overline{D})}\,\||\det J_i|^{-1}J_iJ_i^t\|_{\mathcal{C}^0(\overline{Q_+},S_d(\mathbb{R}))} \lesssim \|a\|_{\mathcal{C}^t(\overline{D})}\,.\end{aligned}$$

We now extend A_i to $\mathbb{R}^d_+$. Since we assumed that $\mathrm{supp}(\chi_i)$ is compact in D_i, we can choose Q_i and $\tilde{Q}_i$ such that $\mathrm{supp}(v_i) \subset Q_i \subset \overline{Q_i} \subset \widetilde{Q}_i \subset \overline{\widetilde{Q}_i} \subset Q$ and consider $\psi \in \mathcal{C}^\infty(\mathbb{R}^d, [0,1])$ such that $\psi = 0$ on Q_i and $\psi = 1$ on $\overline{\widetilde{Q}_i}^c$. We define the extension $\overline{A}_i$ of A_i on $\mathbb{R}^d_+$ by

$$\overline{A}_i(x) := \begin{cases} A_i(x)(1-\psi(x)) + a_{\min}\psi(x)I_d & \text{if } x \in Q_+ \\ a_{\min}\psi(x)I_d & \text{if } x \in Q_+^c \end{cases}$$

Analogously to the case of $\overline{a}$ in Lemma 4.7.4, we can deduce that, for any $y \in \mathbb{R}^d_+$ and $\xi \in \mathbb{R}^d$,

$$\overline{A}_i(y)\xi\cdot\xi \gtrsim a_{\min}|\xi|^2, \quad A_{\max} = \|\overline{A}_i\|_{\mathcal{C}^0(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))} \lesssim a_{\max} \quad \text{and} \quad \|\overline{A}_i(y)\|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))} \lesssim \|a\|_{\mathcal{C}^t(\overline{D})}\,. \tag{4.30}$$

We now define an extension of F_i on $\mathbb{R}^d_+$. Note again that we can choose an open set G_i such that $\mathrm{supp}(F_i) \subset G_i \subset \overline{G}_i \subset Q$ and extend F_i to all of $\mathbb{R}^d_+$ such that $\|F_i\|_{H^{s-1}(\mathbb{R}^d_+)} \lesssim \|F_i\|_{H^{s-1}(Q_+)}$. Finally we continue v_i by 0 on $\mathbb{R}^d_+$, which yields $v_i \in H^1_0(\mathbb{R}^d_+)$. Moreover, since $\overline{A}_i = A_i$ on $\mathrm{supp}(v_i) \subset Q_i$, v_i is then the weak solution on $\mathbb{R}^d_+$ of

$$-\mathrm{div}(\overline{A}_i(x)\nabla v_i(x)) = F_i(x),$$

which enables us to apply Lemma 4.3.3 and to obtain that $v_i \in H^{1+s}(\mathbb{R}^d_+)$ and

$$\|v_i\|_{H^{1+s}(\mathbb{R}^d_+)} \lesssim \frac{A_{\max}}{A_{\min}^2}\Big(|\overline{A}_i|_{\mathcal{C}^t(\overline{\mathbb{R}^d_+},S_d(\mathbb{R}))}\,|v_i|_{H^1(\mathbb{R}^d_+)} + \|F_i\|_{H^{s-1}(\mathbb{R}^d_+)}\Big) + \frac{A_{\max}}{A_{\min}}\|v_i\|_{H^1(\mathbb{R}^d_+)}\,.$$

Recalling that $u_i(x) = v_i(\alpha_i(x))$ for any $x \in D \cap D_i$ and using the bounds in (4.30), as well as the transformation theorem [39, Theorem 6.2.17], we finally get

$$\begin{aligned}\|u_i\|_{H^{1+s}(D)} &\lesssim \frac{a_{\max}}{a_{\min}^2}\Big(\|a\|_{\mathcal{C}^t(\overline{D})}\,\|u\|_{H^1(D\cap D_i)} + \|g_i\|_{H^{s-1}(D\cap D_i)}\Big) + \frac{a_{\max}}{a_{\min}}\|u\|_{H^1(D)} \\ &\lesssim \frac{a_{\max}\,\|a\|_{\mathcal{C}^t(\overline{D})}}{a_{\min}^3}\,\|f\|_{H^{s-1}(D)}\,.\end{aligned}$$

□

The result in Proposition 4.3.1 follows directly from Lemmas 4.7.4 and 4.7.5, if we recall that $u = \sum_{i=0}^{m} u_i$.

Troisième partie

Analyse numérique de l'advection-diffusion d'un soluté dans des milieux aléatoires

Chapitre 5

Numerical analysis of the advection-diffusion of a solute in random media

Abstract : We consider the problem of numerically approximating the solution of the coupling of the flow equation in a random porous medium, with the advection-diffusion equation. More precisely, we present and analyse a numerical method to compute the mean value of the spread of a solute introduced at the initial time, and the mean value of the macro-dispersion, defined as the temporal derivative of the spread. We propose a Monte-Carlo method to deal with the uncertainty, i.e. with the randomness of the permeability field. The flow equation is solved using finite elements. The advection-diffusion equation is seen as a Fokker-Planck equation, and its solution is approximated thanks to a probabilistic particular method. The spread is indeed the expected value of a function of the solution of the corresponding stochastic differential equation, and is computed using an Euler scheme for the stochastic differential equation and a Monte-Carlo method. Error estimates on the mean spread and on the mean dispersion are established, under various assumptions, in particular on the permeability random field.

Keywords : uncertainty quantification, elliptic PDE with random coefficients, advection-diffusion equation, Monte-Carlo method, Euler scheme for SDE.

Résumé : On s'intéresse à l'approximation numérique de la solution du couplage entre l'équation d'écoulement dans un milieu poreux aléatoire et l'équation d'advection-diffusion. Plus précisément, on présente et analyse une méthode numérique pour calculer la valeur moyenne de l'extension d'un soluté introduit au temps initial, et la valeur moyenne de la macro-dispersion, définie comme la dérivée temporelle de l'extension. On propose une méthode de Monte-Carlo pour tenir compte des incertitudes, c'est à dire du caractère aléatoire du champ de perméabilité. L'équation d'écoulement est alors résolue en utilisant des éléments finis. L'équation d'advection-diffusion est vue comme une équation de Fokker-Planck, et sa solution est donc approchée grâce à une méthode particulaire probabiliste. L'extension peut en effet être exprimée comme l'espérance d'une fonction de la solution de l'équation différentielle stochastique correspondante, et est calculée grâce à un schéma d'Euler pour l'équation différentielle stochastique et une méthode de Monte-Carlo. On donne des estimations d'erreur pour l'extension moyenne et la dispersion moyenne, sous différentes hypothèses, en particulier sur le champ de perméabilité.

Mots clés : quantification des incertitudes, EDP elliptique à coefficients aléatoires, équation d'advection-diffusion, méthode de Monte-Carlo, schéma d'Euler pour les EDS.

5.1 Introduction

Numerical modeling is an important key for the management and remediation of groundwater resources. The heterogeneity of natural geological formations has a major impact in the contamination of groundwater by migration of pollutants. In order to account for the limited knowledge of the geological characteristics and for the natural heterogeneity, stochastic models have been developed, see e.g. [16],[17]. The permeability of the porous media is then a random field. Our aim is then to study the migration of a contaminant in steady flow. The flow velocity is computed by solving an elliptic partial differential equation with random coefficients. The solute concentration is then the solution of an advection-diffusion equation, where the flow velocity, which is a random field, appears as a coefficient. The quantities we are interested in are finally the mean value of the spread of the solute, that is to say the mean value of the spatial variance of the solute, and the mean value of the dispersion, which is defined as the derivative of the spread with respect to the time. The determination of the large-scale dispersion coefficients has been widely debated in the last twenty five years, see e.g [5], [14], [18], [28], [65], [67] and [62].

Here we are interested in the case of a lognormal permeability field, which is a widely used model. Moreover we consider the case, physically pertinent, where the correlation length is small and the uncertainty important. Therefore methods based on the approximation of the coefficients in a finite dimensional stochastic space, such as stochastic galerkin methods and stochastic collocation method would be highly expensive, and hence do not seem to be suitable to deal with such cases. Neither seem perturbation methods, since we suppose the uncertainty to be important. As regards the advection-diffusion equation, we focus on the advection-dominated model. Therefore we choose not to consider an Eulerian method, in order to avoid numerical diffusion. The below described method has therefore being proposed and implemented by A.Beaudoin, J.R. de Dreuzy and J.Erhel to compute the mean dispersion in 2D, their numerical results can be found e.g. in [15]. A Monte-Carlo method is used to deal with the uncertainty. The solution of the steady flow equation is computed by using finite elements. The solution of the advection-diffusion equation is approximated using a probabilistic method. We consider the stochastic differential equation associated to this Fokker Planck equation and its solution is approximated with an Euler scheme. A Monte-Carlo method provides finally an approximation to the solution of the Fokker Planck equation. All these steps together lead to an approximation of the mean spread. The mean dispersion is then approximated by the numerical derivative of the computed mean spread.

The aim of this paper is to make the numerical analysis of the above described method. More precisely we furnish a priori error estimates for the approximations of the spread and of the dispersion. We focus on the spread and the dispersion, because of their physical interest, but the result given here are more general. A specifity of this work is to address the coupling of the flow equation with the advection-diffusion equation, whereas most of the existing numerical analysis of methods for uncertainty quantification are limited to the flow equation, see e.g [1], [2]. The main novelty of this work is the use of numerical analysis tools from two differents areas : finite elements method and weak error analysis for SDEs. Moreover, since we estimate the time derivative of the spread, we have to generalize the weak error to estimate the error for time derivatives of averages.

After describing the physical model in section 2, we describe in section 3 the numerical method mentioned above. Section 4 is devoted to the numerical analysis of this method in the case of a random permeability field, supposed to be almost surely periodic, uniformly coercive and bounded under various additional regularity assumptions. We first give preliminary results. The first one is a weak error result for the Euler scheme on a stochastic differential equation with additive noise and $\mathcal{C}^{1,\alpha}$ drift. The second one is a continuity result on the approximation of the solution of a stochastic differential equation using an Euler scheme, with respect to the drift endowed with uniform norm. This result is combined with a classical $W^{1,\infty}$ finite elements error result. After these preliminary results, we give the two main results of this paper, namely error results on the mean spread and on its time derivative, the mean dispersion.

5.2 Physical model

5.2.1 Steady flow equation

We consider an isotropic porous medium, we suppose the porosity to be constant, equal to 1. The domain O is a box in $\mathbb{R}^d$, with $d = 1, 2$, or 3. The heterogeneity of the natural geological formations and the lack of data led us to use a stochastic model. A classical case is the homogeneous lognormal permeability field with a correlation function of the following type :

$$a(\omega, x) = e^{g(\omega,x)},\ x \in O,\ \omega \in \Omega,$$

where g is a gaussian field characterized by its mean m and its covariance function :

$$cov[g](x, y) = \sigma^2 exp\left(-\frac{\|x - y\|^\delta}{l^\delta}\right), \tag{5.1}$$

for some $\delta > 0$. The random parameter is denoted ω. The case of an exponential covariance function corresponds to $\delta = 1$, and furnishes a model which is a reasonably good fit to some field data , see e.g [25] and [42]. Unfortunately, as we will see later, for technical reasons, we are not able to treat this case rigorously.

The variance of the log hydraulic conductivity σ^2 is typically in the interval $[1, 10]$, the correlation length l typically ranges between 0.1m and $100m$, whereas the size of the domain has to be at least hundred times the correlation length l. Classical laws governing the steady flow in porous media without source are mass conservation $\text{div}(v) = 0$ and Darcy law $v = -a\nabla p$, where v is the Darcy velocity and p the hydraulic head. Boundary conditions are homogeneous Dirichlet condition, the hydraulic head on the boundary in denoted by p_0. Finally, the hydraulic head is the solution of the following elliptic PDE with a random coefficient : for almost all ω

$$\text{div}(a(\omega, x)\nabla p(\omega, x)) = 0\ , x \in O. \tag{5.2}$$

this equation is subjected to mixed boundary conditions, and is imposed for almost all $\omega \in \Omega$.
Here, ω is then the parameter describing the randomness of the media. We recall that the Darcy velocity is then defined by

$$v(\omega, x) = -a(\omega, x)\nabla p(\omega, x).$$

5.2.2 Advection-diffusion equation

An inert solute is injected in the porous medium and transported by advection and diffusion. Here we consider only molecular diffusion, assumed homogeneous and isotropic. This type of solute migration is described by the advection-diffusion equation :

$$\frac{\partial c(\omega, x, t)}{\partial t} + v(\omega, x).\nabla_x(c(\omega, x, t)) - D\Delta c(\omega, x, t) = 0, \tag{5.3}$$

where $D > 0$ is the molecular diffusion coefficient, v the Darcy velocity defined previously and c the solute concentration. We consider the case of advection-dominated model, i.e. the case where the Peclet number $Pe = \frac{l\|v\|_{mean}}{D}$ is large (typically ≥ 100). The initial condition at $t = 0$ is the injection of the solute, i.e. $c(t = 0) = \mathbb{1}_R$ where R is a box included in O. Equation (5.3) should be supplemented with boundary conditions on ∂O.

5.2.3 Spread and dispersion

We now define the two quantities we want to compute.
First we introduce the center of mass of the solute distribution :

$$G(\omega, t) = \int_O c(\omega, x, t)xdx.$$

Our aim is then to compute $S(t)$ the mean spread of mass around G, and the mean macro-dispersion $\mathcal{D}(t)$, defined as its time derivative, i.e.

$$S(\omega,t)=\int_O c(\omega,x,t)(x-G(\omega,t))(x-G(\omega,t))^t dx, \quad S(t)=\mathbb{E}_\omega[S(\omega,t)]$$

and

$$\mathcal{D}(\omega,t)=\frac{dS(\omega,t)}{dt}, \quad \mathcal{D}(t)=\mathbb{E}_\omega[\mathcal{D}(\omega,t)].$$

5.3 Description of the numerical method

5.3.1 A Monte-Carlo method to deal with uncertainty

As precised above, we suppose the uncertainty to be large, typically $\sigma^2 \in [1,10]$, therefore perturbation type methods [5], [18], [28], [63], [65], [57], [62] do not seem to be suitable. Moreover, since we suppose l to be small, σ^2 to be large and $cov[g]$ to be only lipschitz, stochastic galerkin and stochastic collocation methods (see e.g. [1], [2], [31], [74] and the references therein) do not seam to be adapted. Namely, in this case, the permeability field a cannot be approximated correctly with a reasonable number of random variable. In particular, the eigenvalues of the Karhunen-Loève development are explicit in this case (see [76]), and we know that the number of term in the truncated Karhunen-Loève development should be much greater than 100, which is not possible on a practical point of view. Therefore we choose to use a Monte-Carlo method to deal with uncertainty.
More precisely, we consider N independent realizations of the permeability field $a(x,\omega_1),...a(x,\omega_N)$. For each i from 1 to N, we compute approximations of the spread $S^i(t)$ and of the dispersion $\mathcal{D}^i(t)$ corresponding to the permeability field a^i as specified below, and we approximate $S(t)$ by $\frac{1}{N}\sum_{i=1}^N S^i(t)$ and $\mathcal{D}(t)$ by $\frac{1}{N}\sum_{i=1}^N \mathcal{D}^i(t)$. For simplicity, the index i will be omitted in the remainder of this section, which is devoted to the description of the numerical method used to compute the solution of a deterministic problem : the computation of spread and dispersion.

5.3.2 Approximation of the flow velocity

The hydraulic head is defined as the solution of the following elliptic partial differential equation :

$$\mathrm{div}(a(x)\nabla p(x))=0,\ x\in O,$$

submitted to mixed boundary conditions.
In what follows, we work with the equivalent PDE with homogeneous mixed boundary conditions and a right hand side f. We define then an approximation $\tilde{p}$ of p in a finite elements space of degree 1, with maximum space mesh h. The velocity v is then approximated by $\tilde{v}=-a\nabla\tilde{p}$. For readibility the maximum mesh size h will be omitted, and the tilde will account for the space discretization approximation in this paper.

5.3.3 A probabilistic particular method

The solute concentration is defined as the solution of (5.3). The domain O is chosen so that a very small amount of the solute reaches the boundary. Therefore, in practice, it is harmless to replace (5.3) by :

$$\left\{\begin{array}{rcll} \frac{\partial c}{\partial t}(x,t)+v(x).\nabla c(x,t)-D\Delta c(x,t) &=& 0, & x\in\mathbb{R}^d \text{ and } t\in[O,T] \\ c(x,0) &=& c_0(x), & x\in\mathbb{R}^d, \end{array}\right.$$

where v is extended to $\mathbb{R}^d$. Since $\mathrm{div}(v)=0$, this is a Fokker-Planck equation. We choose to approximate the solution of this Fokker-Planck equation with a probabilistic particular method, we define the associated stochastic differential equation :

$$\left\{\begin{array}{rcl} dX(t) &=& v(X(t))dt+\sqrt{2D}dW(t) \\ X(0) &=& X_0, \end{array}\right.$$

where X_0 is a random variable with density c_0 with respect to the Lebesgue measure. It is classical that $X(t)$ admits then $c(x,t)dx$ as density. The law of X can be approximated by a Monte-Carlo method. We take M independent realizations of approximations of X using an Euler scheme and the approximated flow velocity $\tilde{v} : X_n^1, ...X_n^M$.

$$\left\{ \begin{array}{rcl} \tilde{X}_n^j(t_{k+1}) & = & \tilde{X}_n^j(t_{k+1}) + \tilde{v}(\tilde{X}_n^j(t_k))\Delta t + \sqrt{2D\Delta t}N_k^j \text{ for } t \in [t_k, t_{k+1}], \\ \tilde{X}_n^j(0) & = & X_0^j, \end{array} \right.$$

where $T = n\Delta t$, $t_k = k\Delta t$ and N^j are independent d-dimensional mean-free gaussian random vector with identity as covariance. Finally we approximate $G(t)$ by $\tilde{G}_n^M(t) = \frac{1}{M}\sum_{j=1}^M \tilde{X}_n^j(t)$, the spread $S(t)$ by

$$\tilde{S}_n^M(t) = \frac{1}{M}\sum_{j=1}^M (\tilde{X}_n^j(t) - \tilde{G}_n^M(t))(\tilde{X}_n^j(t) - \tilde{G}_n^M(t))^t.$$

In order to approximate the dispersion, we introduce a new time step Δs, and approximate $\mathcal{D}(t)$ by

$$\frac{\tilde{S}_n^M(t+\Delta s) - \tilde{S}_n^M(t)}{\Delta s}.$$

Indeed we recall that we have $G(t) = E[X(t)]$, $S(t) = E[(X(t) - G(t))(X(t) - G(t))^t]$, and $\mathcal{D}(t) = \frac{d}{dt}S(t)$.

For more details on a possible numerical implementation, see [15].

5.4 Numerical analysis of the method with additional assumptions

We consider O a box of $\mathbb{R}^d$, and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. For $k \in \mathbb{N}$ and $0 \leq \alpha \leq 1$, we denote by $\mathcal{C}_b^{k,\alpha}$ the space of $\mathcal{C}^k$ functions with bounded derivatives and such that any k-th derivative is α-hölder continuous. For $f \in \mathcal{C}_b^{k,\alpha}$ we introduce associated norm :
$\|f\|_{\mathcal{C}_b^{k,\alpha}} = \max\{\|f\|_\infty, \|f'\|_\infty, ..., \|f^{(k)}\|_\infty, \|f^{(k)}\|_{\mathcal{C}_b^{0,\alpha}}\}$.

The numerical analysis of the above algorithm requires the solution of (5.2) to be sufficiently regular with respect to $x \in O$. Unfortunately, this is not the case for two reasons. First, we are dealing with an elliptic equation on a rectangular domain with mixed boundary conditions. This limits the smoothness of the solution. Also, note that the advection-diffusion equation is set on the full space $\mathbb{R}^d$ and it is not clear how to extend the velocity field on O to $\mathbb{R}^d$. We consider that this is a techincal problem and avoid it by replacing the mixed boundary conditions by periodic boundary conditions, so that the solutions have the smoothness naturally associated to the smoothness of the permeability, and the extension to $\mathbb{R}^d$ is trivial. Another way could be truncate the velocity field close to the boundary and to then extend it by zero. The final solution would not be very different, since in practice the domain O is chosen very large with respect to the box R and a very small amount of the solute reaches the boundary. The solution of (5.2) with mixed boundary conditions being smooth inside the domain, the same analysis as below gives a similar result. We chose the periodic boundary conditions to simplify the presentation.

Second, and this is a much deeper problem, the permeability field associated to an exponential covariance (i.e. $\delta = 1$ in (5.1)), is only $\mathcal{C}^\beta$, for $\beta < 1/2$, yielding a velocity of the same regularity. Our analysis requires at least a $\mathcal{C}^{1,\alpha}$, $\alpha > 0$ regularity, which exludes exponential covariance and we are only able to deal with $\delta > 2 + 2\alpha$ (or $\delta = 2$). Furthermore, as it is often the case in theoritical papers, we assume that the permeability field is uniformly bounded and coercive. This is clearly not the case for a lognormal probability field, but using the argument and methodology of [10] (see also [66] and [26]), we obtain a similar result. Note that this would considerably complicate the proof. To sum up, these considerations lead to the following assumption used in all the results below.

Assumption 5.4.1. *The permeability field $a \in L^\infty(\Omega, \mathcal{C}_b^{1,\alpha}(\mathbb{R}^d))$ for some $0 < \alpha < 1$, such that for almost all ω $a(\omega, .)$ is O periodic and such that for almost all ω, for any $x \in \mathbb{R}^d$,*

$$0 < a_{min} < a(\omega, x) < a_{max} < +\infty,$$

5.4.1 Solution of the flow equation and its approximation using finite elements

We consider then the flow equation :

$$\left\{ \begin{array}{rcll} \mathrm{div}(a(\omega,x)\nabla p(\omega,x)) & = & f, & on\ \Omega\times\mathbb{R}^d, \\ \int_{\partial O} p(\omega,x) & = & 0 & on\ \Omega, \end{array} \right. \tag{5.4}$$

and p is O periodic. The right hand side f takes into account non homogeneous boundary conditions and is assumed to be smooth.

Proposition 5.4.2. *Equation* (5.4) *admits a unique solution* $p \in L^\infty(\Omega, \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d))$.

Proof. For almost all ω, the application of the elliptic regularity theorem [33] yields that the equation admits a unique solution $p(\omega,\cdot) \in \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d)$, with

$$||p(\omega,\cdot)||_{\mathcal{C}_b^{2,\alpha}(\mathbb{R}^d)} \leq C(a(\omega,\cdot))$$

where $C(a(\omega,\cdot))$ is a constant which depends only on the $\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)$ norm of $a(\omega,\cdot)$ and on $a_{\min}(\omega)$ and can therefore be chosen to be uniform with respect to ω. The application $\omega \mapsto p(\omega,\cdot) \in \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d)$ being strongly measurable, the result follows. □

We consider p^h the approximation of p in a finite elements space V_h.

We define the Darcy velocity : $v(\omega,x) = -a(\omega,x)\nabla p(\omega,x)$ and its estimate $v^h(\omega,x) = -a(\omega,x)\nabla p^h(\omega,x)$. Then $v \in L^\infty(\Omega, \mathcal{C}_b^{1,\alpha}(\mathbb{R}^d))$. Hence, since V_h is a finite dimensional space, we also have $v^h \in L^\infty(\Omega\times\mathbb{R}^d)$.

From now on, we make the following assumption on the finite element space.

Assumption 5.4.3. *We suppose that there exists a constant* c_1 *(depending on* v*) such that for any* $h > 0$ *we have* $\|v - v_h\|_{L^\infty(\Omega\times\mathbb{R}^d)} \leq C_1 h|ln(h)|$.

Remark 5.4.4. *This assumption is in particular fullfiled in the case of a finite element space of piecewse linear functions on a regular triangulation. In [21] for instance, this is proved for a* $\mathcal{C}^1$ *permeability field with dirichlet boundary conditions. The proof extends to periodic boundary conditions. Note that we expect that if the permeability field is* $\mathcal{C}^{0,\alpha}$*, a similar estimate holds with* $h^{\alpha-\varepsilon}$ *for* $\varepsilon > 0$*, instead of* $h|ln(h)|$*. We have not found such a result in the litterature but it is possible that it could be proved mixing the argument of [21] and [39], chap. 8 and 9.*

5.4.2 The advection-diffusion equation

We define $c_0(x) = \mathbb{1}_R(x)$, where R is a box included in O, and take $v \in \mathcal{C}_b^{1,\alpha}(\mathbb{R}^d, \mathbb{R}^d)$, we consider then the following advection-diffusion equation :

$$\left\{ \begin{array}{rcll} \frac{\partial c}{\partial t}(x,t) + v(x).\nabla c(x,t) - D\Delta c(x,t) & = & 0, & x \in \mathbb{R}^d \text{ and } t \in [O,T] \\ c(x,0) & = & c_0(x), & x \in \mathbb{R}^d. \end{array} \right. \tag{5.5}$$

We consider also the stochastic differential equation associated to this Fokker Planck equation :

$$\left\{ \begin{array}{rcl} dX(t) & = & v(X(t))dt + \sqrt{2D}dW(t), \\ X(0) & = & X_0, \end{array} \right. \tag{5.6}$$

We suppose that X_0 admits $c_0(x)dx$ as density. We recall the following well known result (see [44] e.g).

Proposition 5.4.5. *The equation* (5.5) *admits a unique solution* $c \in \mathcal{C}(]0,T], \mathcal{C}^2(\mathbb{R}^d))$ *and* $X(t)$ *admits* $c(x,t)dx$ *as density.*

5.4.3 Weak error of time discretization

We give now the first preliminary result. We consider $(\Omega', \mathcal{F}', \mathbb{P}')$ another probability space, whose generic variable is denoted by ξ. Here we give weak error results for the time discretization of a stochastic differential equation with an additive noise and a $\mathcal{C}^{1,\alpha}$ drift.

Let $v \in \mathcal{C}_b^{1,\alpha}(\mathbb{R}^d, \mathbb{R}^d)$ for some $0 < \alpha \leq 1$.
We denote by X^x the solution of the following stochastic differential equation :

$$\left\{ \begin{array}{rcl} dX^x(t) & = & v(X^x(t))dt + \sqrt{2D}dW(t), \\ X^x(0) & = & x. \end{array} \right. \tag{5.7}$$

We denote by X_n^x the numerical approximation of X^x using an Euler scheme as in section 3, where the mesh of the time discretization is $\Delta t = \frac{T}{n}$, and $t_k = k\Delta t$ for $0 \leq k \leq n$. We extend X_n^x to a function defined for all $t \geq 0$ by : :

$$\left\{ \begin{array}{rcl} dX_n^x(t) & = & v(X_n^x(t_k))dt + \sqrt{2D}dW(t), \text{ for } t_k \leq t \leq t_{k+1}, \\ X_n^x(0) & = & x. \end{array} \right. \tag{5.8}$$

We need to deal with a drift less than $\mathcal{C}^2$, this has been studied in [55]. Nevertheless, we give a simple proof in the case of an additive noise for the sake of completness, which moreover yields a better weak convergence order in our particular case (namely $\frac{1+\alpha}{2}$) than the general result of [55] (namely $\frac{1}{2-\alpha}$).

Proposition 5.4.6. *Let $0 < \alpha \leq 1$, $0 < \beta < 1$, $V > 0$ and $\varphi \in \mathcal{C}_b^{2,\beta}(\mathbb{R}^d)$, then there exists a constant C_2 such that for any $x \in \mathbb{R}^d$ and $v \in \mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)$ such that $\|v\|_{\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)} \leq V$, we have*

$$|\mathbb{E}[\varphi(X^x(T)) - \varphi(X_n^x(T))]| \leq C_2(\Delta t)^{\frac{1+\alpha}{2}}.$$

Proof. We denote by $\mathcal{C}_b^{1;2}(\mathbb{R}^d \times [0,T])$ the space of functions of (x,t) which admit derivatives with respect to t and two derivatives with respect to x, all these derivatives being continuous and bounded on $\mathbb{R}^d \times [0,T]$. For $u \in \mathcal{C}_b^{1;2}(\mathbb{R}^d \times [0,T])$, we introduce the natural norm

$$\|u\|_{\mathcal{C}_b^{1;2}(\mathbb{R}^d \times [0,T])} = \max \left\{ \|u\|_\infty, \left\| \frac{\partial u(t,x)}{\partial t} \right\|_\infty, \left\| \frac{\partial u(t,x)}{\partial x} \right\|_\infty, \left\| \frac{\partial^2 u(t,x)}{\partial x^2} \right\|_\infty \right\}.$$

We recall now a classical result whose proof can be found in [52] page 184. We introduce the following Kolmogorov equation associated to the previous SDE :

$$\left\{ \begin{array}{rcl} \frac{\partial u}{\partial t}(t,x) & = & D\Delta u(t,x) + v(x).\nabla u(t,x), \\ u(0,x) & = & \varphi(x). \end{array} \right. \tag{5.9}$$

Then for $0 < \alpha < 1$, $v \in \mathcal{C}_b^{0,\alpha}(\mathbb{R}^d)$ with $\|v\|_{\mathcal{C}_b^{0,\alpha}(\mathbb{R}^d)} \leq V$ and $\varphi \in \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d)$, the equation (5.9) admits a unique solution $u \in \mathcal{C}_b^{1;2+\alpha}(\mathbb{R}^d \times [0,T])$ and $\|u\|_{\mathcal{C}_b^{1;2+\alpha}(\mathbb{R}^d \times [0,T])}$ can be bounded by a constant C_V depending on d, φ, T and depending only on v through V. In particular in our case, since we have $v \in \mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)$ with $\|v\|_{\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)} \leq V$ and $\varphi \in \mathcal{C}_b^{2,\beta}(\mathbb{R}^d)$ then the solution u of (5.9) belongs to $C_b^{1;2}(\mathbb{R}^d \times [0,T])$, with $\|u\|_{\mathcal{C}_b^{1;2}(\mathbb{R}^d \times [0,T])} \leq C_V$

Applying Itô formula yields classically $u(t,x) = \mathbb{E}[\varphi(X^x(t))]$. Then the weak error

$$\begin{array}{rcl} E & = & \mathbb{E}[\varphi(X^x(T))] - \mathbb{E}[\varphi(X_n^x(T))] \\ & = & u(T,x) - \mathbb{E}[u(0, X_n^x(T))] \end{array}$$

can be split into $E = \sum_{i=0}^{n-1} E_i$, where :

$$\begin{array}{rcl} E_i & = & \mathbb{E}[u(T - t_i, X_n^x(t_i))] - \mathbb{E}[u(T - t_{i+1}, X_n^x(t_{i+1}))] \\ & = & \mathbb{E}[u(T - t_{i+1}, X^{X_n^x(t_i)}(\Delta t))] - \mathbb{E}[u(T - t_{i+1}, X_n^{X_n^x(t_i)}(\Delta t))] \\ & = & \mathbb{E}[\mathbb{E}[u(T - t_{i+1}, X^{X_n^x(t_i)}(\Delta t)) - u(T - t_{i+1}, X_n^{X_n^x(t_i)}(\Delta t))|X_n^x(t_i)]] \\ & = & \mathbb{E}[e_i(X_n^x(t_i))], \end{array}$$

where $e_i(y) = \mathbb{E}[u(T - t_{i+1}, X^y(\Delta t)) - u(T - t_{i+1}, X_n^y(\Delta t))]$, by using the Markov property of the solution X and of the discretized solution X_n. Using a Taylor expansion of u at order two with respect to x, we get :

$$\begin{aligned} |e_i(y)| &\leq |\mathbb{E}[D_x u(T - t_{i+1}, X^y(\Delta t)).(X_n^y(\Delta t) - X^y(\Delta t))]| \\ &+ \|D_x^2 u\|_\infty \mathbb{E}\left[|X^y(\Delta t) - X_n^y(\Delta t)|^2\right]. \end{aligned}$$

We first bound the second term, using Cauchy-Schwarz inequality :

$$\begin{aligned} \mathbb{E}\left[|X^y(\Delta t) - X_n^y(\Delta t)|^2\right] &\leq \Delta t \mathbb{E}\left[\int_0^{\Delta t} |v(X^y(s)) - v(y)|^2 ds\right] \\ &\leq \|Dv\|_\infty^2 \Delta t \mathbb{E}\left[\int_0^{\Delta t} |X^y(s) - y|^2 ds\right]. \end{aligned}$$

The integrand can be bounded as follows :

$$\begin{aligned} |X^y(s) - y| &= \left|\int_0^s v(X^y(t))dt + \sqrt{2D}W(t)\right| \\ &\leq \|v\|_\infty s + \sqrt{2D}|W(s)|, \end{aligned}$$

which yields the following bound for the second term :

$$\begin{aligned} \mathbb{E}\left[|X^y(\Delta t) - X_n^y(\Delta t)|^2\right] &\leq \|Dv\|_\infty^2 \Delta t \int_0^{\Delta t} \mathbb{E}\left[\left(\|v\|_\infty \Delta t + \sqrt{2D\Delta t}\left|\frac{W(s)}{\sqrt{s}}\right|\right)^2 ds\right] \\ &\leq \|Dv\|_\infty^2 (\Delta t)^2 \int_0^{\Delta t} \mathbb{E}\left[(\|v\|_\infty \sqrt{\Delta t} + \sqrt{2D}\,|W(1)|)^2 ds\right] \\ &\leq 2(\sqrt{2D} + \|v\|_\infty^2)\|Dv\|_\infty^2 (\Delta t)^3. \end{aligned}$$

We now bound the first term, since $v \in \mathcal{C}^{1,\alpha}(\mathbb{R}^d)$, for any $x, y \in \mathbb{R}^d$

$$\begin{aligned} |v(x) - v(y) - Dv(x).(y - x)| &= \left|\int_0^1 (Dv(x + t(y - x)) - Dv(x)).(y - x))dt\right| \\ &\leq \|v\|_{\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)} |y - x|^{\alpha+1}. \end{aligned}$$

Therefore, using a Taylor expansions of v and $D_x u$, we get

$$\begin{aligned} & |\mathbb{E}\left[D_x u(T - t_{i+1}, X^y(\Delta t)).(X_n^y(\Delta t) - X^y(\Delta t))\right]| \\ = & \left|\mathbb{E}\left[D_x u(T - t_{i+1}, X^y(\Delta t)).\int_0^{\Delta t} (v(X^y(u)) - v(y))du\right]\right| \\ \leq & \left|\mathbb{E}\left[D_x u(T - t_{i+1}, X^y(\Delta t)).\int_0^{\Delta t} Dv(y).(X^y(s) - y)ds\right]\right| \\ + & \|v\|_{\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)} \|Du\|_\infty \mathbb{E}\left[\int_0^{\Delta t} |X^y(s) - y|^{\alpha+1} ds\right] \\ \leq & \left|\mathbb{E}\left[(D_x u(T - t_{i+1}, X^y(\Delta t)) - D_x u(T - t_{i+1}, y)).\int_0^{\Delta t} Dv(y).(X^y(s) - y)ds\right]\right| \\ + & |D_x u(T - t_{i+1}, y).\int_0^{\Delta t} Dv(y).\mathbb{E}\left[X^y(s) - y\right] ds| \\ + & \|v\|_{\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)} \|Du\|_\infty \mathbb{E}\left[\int_0^{\Delta t} |(\|v\|_\infty s + \sqrt{2D}|W(s)|)|^{\alpha+1} ds\right] \end{aligned}$$

Hence, we have

$$\begin{aligned}
& |\mathbb{E}\left[D_x u(T-t_{i+1}, X^y(\Delta t)).(X_n^y(\Delta t) - X^y(\Delta t))\right]| \\
\leq\ & \|D_x^2 u\|_\infty \|Dv\|_\infty \mathbb{E}\left[|X^y(\Delta t) - y| \int_0^{\Delta t} |X^y(s) - y| ds\right] \\
+\ & \|D_x u\|_\infty \|Dv\|_\infty \int_0^{\Delta t} |\mathbb{E}\left[X^y(u) - y\right]| du \\
+\ & \|v\|_{\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)} \|Du\|_\infty \int_0^{\Delta t} \mathbb{E}\left[(\|v\|_\infty \Delta t + \sqrt{2D\Delta t}|W(1)|)^{\alpha+1}\right] ds \\
\leq\ & \|D_x^2 u\|_\infty \|Dv\|_\infty \left(\mathbb{E}[(X^y(\Delta t) - y)^2]\right)^{1/2} \left(\mathbb{E}\left[\left(\int_0^{\Delta t} |X^y(s) - y| ds\right)^2\right]\right)^{1/2} \\
+\ & \|D_x u\|_\infty \|Dv\|_\infty \int_0^{\Delta t} \|v\|_\infty u du \\
+\ & \|v\|_{\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d)} \|Du\|_\infty \int_0^{\Delta t} \mathbb{E}\left[(\|v\|_\infty \Delta t + \sqrt{2D\Delta t}|W(1)|)^{\alpha+1}\right] ds \\
\leq\ & C((\Delta t)^2 + (\Delta t)^{1+\frac{\alpha+1}{2}}).
\end{aligned}$$

The constant C which appears in this bound depends only on V, φ, d and T. These two estimates lead to the following bound for e_i :

$$|e_i(y)| \leq C(\Delta t)^{1+\frac{1+\alpha}{2}}.$$

The final result follows by taking the sum over i of these inequalities, recalling that $n\Delta t = T$. □

Remark 5.4.7. *If $\alpha = 0$, the result on the weak order holds as a consequence of the result on the strong order, namely in this case, the drift is lipschitz and therefore it is classical that the strong order of the Euler scheme is 1/2 (see [56] and [37]).*

Remark 5.4.8. *If the drift only belongs to $\mathcal{C}_b^{0,\alpha}(\mathbb{R}^d)$, we have a similar result.*

Let $0 < \alpha \leq 1$, $0 < \beta < 1$, $\varphi \in \mathcal{C}_b^{2,\beta}(\mathbb{R}^d)$, and $v \in \mathcal{C}_b^{0,\alpha}$ such that $\|v\|_{\mathcal{C}^{0,\alpha}} \leq V$, then there exists a constant C depending only on φ, α, d, T and V such that

$$|\mathbb{E}[\varphi(X^x(T)) - \varphi(X_n^x(T))]| \leq C(\Delta t)^{\frac{\alpha}{2}}.$$

5.4.4 Space discretization error on the solution of the SDE

We give now the second preliminary result. We consider the Euler scheme with an exact velocity v, where $v \in \mathcal{C}_b^1(\mathbb{R}^d)$:

$$\left\{\begin{array}{rcl} dX_n(t) & = & v(X_n(t_k))dt + \sqrt{2D}dW(t), \text{ for } t \in [t_k, t_{k+1}], \\ X_n(0) & = & X_0, \end{array}\right.$$

and the Euler scheme with an approximated velocity $\tilde{v}$, where $\tilde{v} \in L^\infty(\mathbb{R}^d)$:

$$\left\{\begin{array}{rcl} d\tilde{X}_n(t) & = & \tilde{v}(\tilde{X}_n(t_k))dt + \sqrt{2D}dW(t), \text{ for } t \in [t_k, t_{k+1}], \\ \tilde{X}_n(0) & = & X_0. \end{array}\right.$$

We give here a bound of the error between these two stochastic processes.

Proposition 5.4.9. *For any T and V, there exists a constant C_3 such that for any v such that $\|Dv\|_\infty \leq V$ we have almost surely*

$$\sup_{t\in[0,T]} |\tilde{X}_n(t) - X_n(t)| \leq C_3 ||v - \tilde{v}||_{L^\infty(\mathbb{R}^d)}.$$

Proof. For any $0 \leq k \leq n-1$ we define $u_k = \sup_{t\in[t_k,t_{k+1}]} |\tilde{X}_n(t) - X_n(t)|$, and we also define $u_{-1} = t_{-1} = 0$. Take $-1 \leq k \leq n-1$, $t \in [t_k, t_{k+1}]$, the difference between the two Euler schemes is then

$$\begin{aligned}
\tilde{X}_n(t) - X_n(t) &= \tilde{X}_n(t_k) - X_n(t_k) + (\tilde{v}(\tilde{X}_n(t_k)) - v(X_n(t_k)))(t - t_k), \\
|\tilde{X}_n(t) - X_n(t)| &\leq |\tilde{X}_n(t_k) - X_n(t_k)| \\
&+ (|\tilde{v}(\tilde{X}_n(t_k)) - v(\tilde{X}_n(t_k))| + |v(\tilde{X}_n(t_k)) - v(X_n(t_k))|)\Delta t
\end{aligned}$$

$$\begin{aligned}
u_{k+1} &\leq u_k + (||v - \tilde{v}||_{L^\infty} + V u_k)\Delta t \\
u_{k+1} &\leq (1 + V\Delta t)u_k + \Delta t ||v - \tilde{v}||_{L^\infty}.
\end{aligned}$$

Hence, the discrete Gronwall lemma implies that for any $0 \leq k \leq N-1$:

$$\begin{aligned}
u_k &\leq \frac{(1 + \Delta t V)^{k+1}}{V} ||v - \tilde{v}||_{L^\infty} \\
&\leq \frac{e^{VT}}{V} ||v - \tilde{v}||_{L^\infty}.
\end{aligned}$$

□

5.4.5 Total error on the spread

We recall that $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega', \mathcal{F}', \mathbb{P}')$ are two probability spaces, with generic variables $\omega \in \Omega$ and $\xi \in \Omega'$. Let $\varphi \in \mathcal{C}_b^3(\mathbb{R}^d, \mathbb{R}^p)$ and $\psi \in \mathcal{C}_b^1(\mathbb{R}^p, \mathbb{R}^q)$ for some $p, q \in \mathbb{N}^*$. Take a constant κ such that φ, ψ and their derivatives are bounded by κ. For almost all $\omega \in \Omega$ we define $X(\omega, \xi, t)$ as the solution of the following stochastic differential equation :

$$\begin{cases} dX(\omega,\xi,t) = v(\omega, X(\omega,\xi,t))dt + \sqrt{2D}dW(\xi,t), \ x \in \mathbb{R}^d, t \geq 0, \\ X(\omega,\xi,0) = X_0(\xi), \end{cases} \tag{5.10}$$

where v is defined as in subsection 4.1, W is a d-dimensional brownian motion on $(\Omega', \mathcal{F}', \mathbb{P}')$ and X_0 admits c_0 as density, as defined in section 4.2. Then we define for any $1 \leq i \leq N$, $1 \leq j \leq M$ and almost all ω its approximations $\tilde{X}_n^{i,j}(\omega, \xi, t)$ by :

$$\begin{cases} d\tilde{X}_n^{i,j}(\omega,\xi,t) &= \tilde{v}^i(\omega, \tilde{X}_n^{i,j}(\omega,\xi,t_k))dt + \sqrt{2D}dW^{i,j}(\xi,t), \text{ for } t \in [t_k, t_{k+1}] \\ \tilde{X}_n^{i,j}(\omega,\xi,0) &= X_0^{i,j}(\xi), \end{cases} \tag{5.11}$$

where $\tilde{v}^i$ is the finite element approximation of v^i as defined in subsection 4.1. We now define the error :

Definition 5.4.10. *For almost all $\omega \in \Omega$, $\xi \in \Omega'$, we define :*

$$Er(\omega,\xi) = \mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))])] - \frac{1}{N}\sum_{i=1}^N \psi\left(\frac{1}{M}\sum_{j=1}^M \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))\right).$$

Theorem 5.4.11. *There exists a constant C such that*

$$\|Er(\omega,\xi)\|_{L^2_{\omega,\xi}} \leq C\left((\Delta t)^{\frac{1+\alpha}{2}} + h|ln(h)| + \frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}\right).$$

Remark 5.4.12. *We expect that for a less regular permeability field, that is assumed to be only $\mathcal{C}^{0,\alpha}$, a similar result holds with the right hand side replaced by*

$$\Delta t^{\frac{\alpha}{2}} + h^{\alpha-\varepsilon} + \frac{1}{\sqrt{M}} + \frac{1}{\sqrt{N}}.$$

Reminding Remark 5.4.4 and Remark 5.4.8, the finite element error and the bound error due to time discretization generalize. The major difficulty is the generalization of Proposition 5.4.9. Note that Proposition 5.4.9 gives a strong error, i.e. for each $\xi \in \Omega'$. In the proof below, only a weak error is needed. We believe that such a weak error of the form

$$|\mathbb{E}_\xi[\varphi(\tilde{X}_n(t))] - \mathbb{E}_\xi[\varphi(X_n(t))]| \leq C\|v - \tilde{v}\|_{L^\infty(\Omega\times\mathbb{R}^d)}$$

is true for a permeability field in $\mathcal{C}^{0,\alpha}$, but we are not able to prove this.

Remark 5.4.13. *An estimate of the error on the spread follows from the cases where $\varphi(x) = xx^t$, $\psi(x) = x$ and $\varphi(x) = x$, $\psi(x) = xx^t$. For simplicity, we treat only the case where φ and ψ are bounded with bounded derivatives. The result can however be generalized to the case where φ and ψ are respectively $\mathcal{C}^3$ and $\mathcal{C}^1$ with at most polynomial growth.*

Proof. Take $V = \|v\|_{L^\infty(\Omega,\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d))}$.
We split this error into three terms :

$$Er(\omega,\xi) = Er1 + Er2(\omega) + Er3(\omega,\xi).$$

Where we define :

$$\begin{aligned}
Er1 &= \mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))])] - \mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))])]\\
Er2(\omega) &= \mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))])] - \frac{1}{N}\sum_{i=1}^N \psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))])\\
Er3(\omega,\xi) &= \frac{1}{N}\sum_{i=1}^N \left(\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))]) - \psi\left(\frac{1}{M}\sum_{j=1}^M \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))\right)\right).
\end{aligned}$$

The first error term $Er1$ takes into account the space discretization error and the time discretization error, and can thus be split into two terms : for almost all ω we have

$$\begin{aligned}
&|\mathbb{E}_\xi[\varphi(X(\omega,\xi,T)] - \mathbb{E}_\xi[\varphi(\tilde{X}_n(\omega,\xi,T))]|\\
&\leq |\mathbb{E}_\xi[\varphi(X(\omega,\xi,T)] - \mathbb{E}_\xi[\varphi(X_n(\omega,\xi,T))]|\\
&+|\mathbb{E}_\xi[\varphi(X_n(\omega,\xi,T)] - \mathbb{E}_\xi[\varphi(\tilde{X}_n(\omega,\xi,T))]|\\
&\leq C_2(\Delta t)^{\frac{\alpha+1}{2}} + C_3\|v - \tilde{v}\|_{L^\infty(\Omega\times\mathbb{R}^d)}\\
&\leq C_2(\Delta t)^{\frac{\alpha+1}{2}} + C_4 h|ln(h)|,
\end{aligned}$$

where we have used Proposition 5.4.6 to bound the first term and Propositions 5.4.9 and 5.4.3 to bound the second term. The constant V which appears in Propositions 5.4.9 and 5.4.6 is then the constant V defined above, i.e. $V = \|v\|_{L^\infty(\Omega,\mathcal{C}_b^{1,\alpha}(\mathbb{R}^d))}$. This inequality holds for almost all ω, then by taking the expected value of the image by ψ we obtain :

$$\begin{aligned}
|Er1| &\leq \mathbb{E}[\|\psi'\|_\infty(C_2(\Delta t)^{\frac{\alpha+1}{2}} + C_4 h|ln(h)|)]\\
&\leq \kappa(C_2(\Delta t)^{\frac{\alpha+1}{2}} + C_4 h|ln(h)|).
\end{aligned}$$

The random variables $(Y_i)_{1\leq i\leq N}$ defined by $Y_i(\omega) = \psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))]$ being independent, identically distributed and belonging to $L^2(\Omega)$, we have :

$$\begin{aligned}
\|Er2(\omega)\|_{L^2_\omega} &\leq \frac{\|Y_i - \mathbb{E}[Y_i]\|_{L^2_\omega}}{\sqrt{N}}\\
&\leq \frac{2\kappa}{\sqrt{N}}.
\end{aligned}$$

Indeed, for almost all ω, we have $|Y_i(\omega)| \leq \|\psi\|_\infty$. Analogously, for any $1 \leq i \leq N$ and almost all ω, the random variables $(Z_j)_{1\leq j\leq M}$ defined by $Z_j(\xi) = \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))$ are independent, identically distributed $L^2(\Omega')$ random variables, therefore :

$$\begin{aligned}\left\|\mathbb{E}[Z_j] - \frac{1}{M}\sum_{j=1}^{M} Z_j(\xi)\right\|_{L^2_\xi} &\leq \frac{\|Z_j - \mathbb{E}[Z_j]\|_{L^2_\xi}}{\sqrt{M}} \\ &\leq \frac{2\kappa}{\sqrt{M}}.\end{aligned}$$

For all $1 \leq i \leq N$ and almost all ω,

$$\begin{aligned}&\left|\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))]) - \psi\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))\right)\right| \\ &\leq \|\psi'\|_\infty \left|\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))] - \frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))\right|,\end{aligned}$$

thus

$$\begin{aligned}&\left\|\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))]) - \psi\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))\right)\right\|_{L^2_\xi} \\ &\leq \frac{2\kappa^2}{\sqrt{M}}.\end{aligned}$$

This bound holds for any $1 \leq i \leq N$ and almost all ω, therefore taking the sum over i and the L^2_ω norm yields finally the following bound for $Er3$:

$$\|Er3(\omega,\xi)\|_{L^2_\omega L^2_\xi} \leq \frac{2\kappa^2}{\sqrt{M}}.$$

□

5.4.6 Total error on the dispersion

We recall that $(\Omega,\mathcal{F},\mathbb{P})$ and $(\Omega',\mathcal{F}',\mathbb{P}')$ are two probability spaces. Let $\varphi \in \mathcal{C}_b^5(\mathbb{R}^d,\mathbb{R}^p)$ and $\psi \in \mathcal{C}_b^2(\mathbb{R}^p,\mathbb{R}^q)$ for some $p,q \in \mathbb{N}^*$. Take a constant κ such that φ,ψ and their derivatives are bounded by κ. $X(\omega,\xi,t)$ and its approximations $\tilde{X}_n^{i,j}(\omega,\xi,t)$ are defined as previously by respectively (5.10) and (5.11).

In this section we give a bound for the below defined error.

Definition 5.4.14. *for almost all $\omega \in \Omega, \xi \in \Omega'$, we define*

$$\begin{aligned}E(\omega,\xi) &= \frac{d}{dt}\mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))])] \\ &- \frac{1}{N}\sum_{i=1}^{N}\frac{\psi\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s)\right) - \psi\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T)\right)}{\Delta s},\end{aligned}$$

where we have naturally defined

$$\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s) = \tilde{X}_n^{i,j}(T) + \tilde{v}(\tilde{X}_n^{i,j}(T))\Delta s + \sqrt{2D\Delta s}(W^{i,j}(T+\Delta s) - W^{i,j}(T)).$$

To bound the error, we make the following additional assumption :

Assumption 5.4.15. *$a \in L^\infty(\Omega, \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d))$ for some $0 < \alpha < 1$.*

Theorem 5.4.16. *There exists a constant C such that*

$$\|E(\omega,\xi)\|_{L^2_{\omega,\xi}} \leq C\left(\Delta t + \Delta s + h|ln(h)| + \frac{1}{\sqrt{N}} + \frac{1}{\sqrt{M\Delta s}}\right).$$

Remark 5.4.17. *The last term in the right hand side (namely $\frac{1}{\sqrt{M\Delta s}}$) is due to the fact that we approximate the derivative of the expected value of a function of the solution of a SDE through a Monte-Carlo method (it comes from the term $E4$ in the following proof), and is optimal (as we can see with the case of a constant drift).*

Remark 5.4.18. *An estimate of the error on the dispersion follows from the cases where $\varphi(x) = xx^t$, $\psi(x) = x$ and $\varphi(x) = x$, $\psi(x) = xx^t$. For simplicity, we treat only the case where φ and ψ are bounded with bounded derivatives. The result can however be generalized to the case where φ and ψ are respectively $\mathcal{C}^5$ and $\mathcal{C}^2$ with growth at most polynomial.*

Proof. We split this error into four terms :

$$E(\omega,\xi) = E1 + E2 + E3(\omega) + E4(\omega,\xi).$$

These four terms are defined by :

$$\begin{aligned}
E1 &= \frac{d}{dt}\mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))])] \\
&- \mathbb{E}_\omega\left[D\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T)]).\frac{\mathbb{E}_\xi[\varphi(X(\omega,\xi,T+\Delta s))] - \mathbb{E}_\xi[\varphi(X(\omega,\xi,T))]}{\Delta s})\right], \\
E2 &= \mathbb{E}_\omega\left[D\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))].\frac{\mathbb{E}_\xi[\varphi(X(\omega,\xi,T+\Delta s))] - \mathbb{E}_\xi[\varphi(X(\omega,\xi,T))]}{\Delta s}\right] \\
&- \frac{\mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s))]] - \mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))]]}{\Delta s}, \\
E3(\omega) &= \frac{\mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s))]] - \mathbb{E}_\omega[\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))]]}{\Delta s} \\
&- \frac{\frac{1}{N}\sum_{i=1}^N \psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s))]) - \psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))])}{\Delta s}, \\
E4(\omega,\xi) &= \frac{\frac{1}{N}\sum_{i=1}^N \psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s))]) - \psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))])}{\Delta s} \\
&- \frac{\frac{1}{N}\sum_{i=1}^N \psi\left(\frac{1}{M}\sum_{j=1}^M \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s))\right) - \psi\left(\frac{1}{M}\sum_{j=1}^M \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))\right)}{\Delta s}.
\end{aligned}$$

E1 The first term $E1$ accounts for the approximation of the time derivative. To bound this term, we first prove the following lemma :

Lemma 5.4.19. *Take a constant V, $0 < \alpha < 1$, then there exists a constant C_V such that for any $v \in \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d)$ with $\|v\|_{\mathcal{C}_b^{2,\alpha}(\mathbb{R}^d)} \leq V$, the solution Y^x of the stochastic differential equation (5.8) is such that for any $x \in \mathbb{R}^d$ we have*

$$\left|\frac{\mathbb{E}[\varphi(Y^x(T+\Delta s))] - \mathbb{E}[\varphi(Y^x(T))]}{\Delta s} - \frac{d}{dt}\mathbb{E}[\varphi(Y^x(t))]\right| \leq C_V \Delta s.$$

Proof. As previously we define u as the solution of (5.9), then we have $u \in \mathcal{C}_b^{1;2}(\mathbb{R}^d \times [0,2T])$ as previously, and thanks to the additionnal assumptions, applying the result of [52] to $\frac{\partial u}{\partial t}$, we have $\frac{\partial u}{\partial t} \in \mathcal{C}_b^{1;2}(\mathbb{R}^d \times [0,2T])$ with $\|\frac{\partial u}{\partial t}\|_{\mathcal{C}^{1;2}(\mathbb{R}^d \times [0,2T])} \leq C_V$. Therefore, we have for any $x \in \mathbb{R}^d$

$$\begin{aligned}
&\left|\frac{\mathbb{E}[\varphi(Y^x(T+\Delta s))] - \mathbb{E}[\varphi(Y^x(T))]}{\Delta s} - \frac{d}{dt}\mathbb{E}[\varphi(Y^x(t))]\right| \\
&= \left|\frac{u(T+\Delta s, x) - u(T,x)}{\Delta s} - \frac{\partial}{\partial t}u(T,x)\right| \\
&\leq \Delta s \int_0^1 (1-t)\left|\frac{\partial^2 u}{\partial t^2}(T+u\Delta s, x)\right| du \\
&\leq \Delta s \sup_{x\in\mathbb{R}^d, s\in[T,T+\Delta s]}\left|\frac{\partial^2 u}{\partial t^2}(s,x)\right| \\
&\leq C_V \Delta s.
\end{aligned}$$

□

We can now bound $E1$, first we notice that since we have supposed that $a \in L^\infty(\Omega, \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d))$, we have $v \in L^\infty(\Omega, \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d))$ by [33], as in the proof of Proposition 5.4.2. Therefore, setting $V = \|v\|_{L^\infty(\Omega, \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d))}$, then we can apply the previous lemma for almost all ω, and since the bound $C_V\Delta s$ is independent from the deterministic initial condition, the Markov property yields that for almost all ω :

$$\begin{aligned}
&\left|\frac{d}{dt}\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))] - \frac{\mathbb{E}_\xi[\varphi(X(\omega,\xi,T+\Delta s))] - \mathbb{E}_\xi[\varphi(X(\omega,\xi,T))]}{\Delta s}\right| \\
&\leq C_V \Delta s.
\end{aligned}$$

$E1$ can be rewritten as

$$\begin{aligned}
E1 &= \mathbb{E}_\omega[D\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))]).\frac{d}{dt}(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T))]) \\
&- \mathbb{E}_\omega\left[D\psi(\mathbb{E}_\xi[\varphi(X(\omega,\xi,T)]).\frac{\mathbb{E}_\xi[\varphi(X(\omega,\xi,T+\Delta s))] - \mathbb{E}_\xi[\varphi(X(\omega,\xi,T))]}{\Delta s}\right]
\end{aligned}$$

Using the previous bound, bounding $\|\psi'\|_\infty$ by κ and taking the expected value with respect to ω yields the following bound for $E1$:

$$|E1| \leq \kappa C_V \Delta s.$$

E2 The term $E2$ contains the space and time discretizations errors. We define once again $V = \|v\|_{L^\infty(\Omega, \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d))}$. The estimate of $E2$ will follow from an estimate of :

$$\begin{aligned}
A &= \frac{\mathbb{E}_\xi[\varphi(\tilde{X}_n(\omega,\xi,T+\Delta s))] - \mathbb{E}_\xi[\varphi(\tilde{X}_n(\omega,\xi,T))]}{\Delta s} \\
&- \frac{\mathbb{E}_\xi[\varphi(X(\omega,\xi,T+\Delta t))] - \mathbb{E}_\xi[\varphi(X(\omega,\xi,T))]}{\Delta s}.
\end{aligned}$$

Actually A corresponds to the case where $\psi(x) = x$. To simplify the variable ω will be fixed and omitted in this paragraph, except in the final bound, and all bounds are uniform with respect to ω, the index i,j will be also omitted, since the random processes are identically distributed and the variable ξ will be also omitted, all expected values are implicitly taken with respect to ξ.
We split the error term A into a time discretization error B and a space discretization error C, i.e. $A = B + C$ where

$$B = \frac{\mathbb{E}[\varphi(X_n(T+\Delta s))] - \mathbb{E}[\varphi(X_n(T))]}{\Delta s} - \frac{\mathbb{E}[\varphi(X(T+\Delta t))] - \mathbb{E}[\varphi(X(T))]}{\Delta s},$$

$$C = \frac{\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))] - \mathbb{E}[\varphi(\tilde{X}_n(T))]}{\Delta s} - \frac{\mathbb{E}[\varphi(X_n(T+\Delta t))] - \mathbb{E}[\varphi(X_n(T))]}{\Delta s}.$$

We first bound the time discretization error B. For any $\phi \in \mathcal{C}_b(\mathbb{R}^d)$ we introduce the following notations : $P^t(\phi)(x) = \mathbb{E}[\phi(X^x(t))]$ and $P_n^t(\phi)(x) = \mathbb{E}[\phi(X_n^x(t))]$ where X^x and X_n^x are defined by (5.7) and (5.8) respectively. For $v \in \mathcal{C}_b^2(\mathbb{R}_d)$, Proposition 5.4.6 in the case $\alpha = 1$ can be reformulated as follows : for any $\phi \in \mathcal{C}_b^{2,\beta}(\mathbb{R}^d)$ for some $0 < \beta < 1$ we have, recalling that $T = n\Delta t$,

$$|P^{\Delta s}(\phi)(x) - P_n^{\Delta s}(\phi)(x)| \leq C_2(\Delta s)^2 \text{ and } |P^T(\phi)(x) - P_n^T(\phi)(x)| \leq C_2\Delta t \tag{5.12}$$

where the constant C_2 depends only on v through V. The following inequality clearly hold true : for any t, and $\phi \in \mathcal{C}_b$,
$\|P^t(\phi)\|_\infty \leq \|\phi\|_\infty$ and $\|P_n^t(\phi)\|_\infty \leq \|\phi\|_\infty$.
We can now bound B, using these notations and inequalities. The Markov property of X and X_n yields :

$$\begin{aligned} B &= \mathbb{E}\left[P_n^T\left(\frac{P_n^{\Delta s} - Id}{\Delta s}\right)\varphi(X_0) - P^T\left(\frac{P^{\Delta s} - Id}{\Delta s}\right)\varphi(X_0)\right] \\ &= \mathbb{E}\left[P_n^T\left(\frac{P_n^{\Delta s} - P^{\Delta s}}{\Delta s}\right)\varphi(X_0) + (P_n^T - P^T)\left(\frac{P^{\Delta s} - Id}{\Delta s}\right)\varphi(X_0)\right]. \end{aligned}$$

For any $x \in \mathbb{R}^d$,

$$\left|\left(\frac{P_n^{\Delta s} - P^{\Delta s}}{\Delta s}\right)\varphi(x)\right| \leq C_2\Delta s,$$

whence for any $x \in \mathbb{R}^d$,

$$\left|P_n^T\left(\frac{P_n^{\Delta s} - P^{\Delta s}}{\Delta s}\right)\varphi(x)\right| \leq C_2\Delta s,$$

and

$$\left|\mathbb{E}\left[P_n^T\left(\frac{P_n^{\Delta s} - P^{\Delta s}}{\Delta s}\right)\varphi(X_0)\right]\right| \leq C_2\Delta s.$$

For any $x \in \mathbb{R}^d$,

$$\frac{P^{\Delta s} - Id}{\Delta s}\varphi(x) = \frac{u(\Delta s, x) - u(0,x)}{\Delta s},$$

therefore

$$\left|\frac{P^{\Delta s} - Id}{\Delta s}\varphi(x) - \frac{\partial u}{\partial t}(0,x)\right| \leq \Delta s\left|\sup_{s\in[0,T],x\in\mathbb{R}^d}\frac{\partial^2 u}{\partial t^2}\right|,$$

where u is still defined by (5.9). And finally, for any $x \in \mathbb{R}^d$,

$$\begin{aligned} &\left|(P_n^T - P^T)\left(\frac{P^{\Delta s} - Id}{\Delta s}\right)\varphi(x)\right| \\ &\leq \left|(P_n^T - P^T)\left(\frac{\partial u}{\partial t}(0,.)\right)(x)\right| \\ &+2\Delta s \sup_{s\in[0,T],x\in\mathbb{R}^d}\left|\frac{\partial^2 u}{\partial t^2}(s,x)\right| \\ &\leq C_2\Delta t + 2\Delta s C_V, \end{aligned}$$

by applying (5.12) with $\phi = \frac{\partial u}{\partial t}(0,.) = D\Delta\varphi + v.\nabla\varphi \in \mathcal{C}_b^{2,\alpha}(\mathbb{R}^d)$.
From the two previous bound, we deduce a bound for B :

$$|B| \leq C_2\Delta t + (2C_V + C_2)\Delta s.$$

We now give a bound for the term of space discretization error C.
We first introduce the notation $\tilde{P}_n^t(\varphi)(x) = \mathbb{E}[\varphi(\tilde{X}_n(t))]$ and use the Markov property of X_n and $\tilde{X}_n$:

$$\begin{aligned} C &= \frac{\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))] - \mathbb{E}[\varphi(\tilde{X}_n(T))]}{\Delta s} - \frac{\mathbb{E}[\varphi(X_n(T+\Delta t))] - \mathbb{E}[\varphi(X_n(T))]}{\Delta s} \\ &= \mathbb{E}\left[\frac{(\tilde{P}_n^{T+\Delta s}(\varphi)(X_0) - \tilde{P}_n^T(\varphi)(X_0)) - (P_n^{T+\Delta s}(\varphi)(X_0) - P_n^T(\varphi)(X_0))}{\Delta s}\right]. \end{aligned}$$

The Markov property enables us to treat the case of a deterministic initial condition. We have now to bound the below defined function $F(x)$, uniformly with respect to x :

$$F(x) = \frac{\mathbb{E}[\varphi(\tilde{X}_n^x(T+\Delta s))] - \mathbb{E}[\varphi(\tilde{X}_n^x(T))]}{\Delta s} - \frac{\mathbb{E}[\varphi(X_n^x(T+\Delta t))] - \mathbb{E}[\varphi(X_n^x(T))]}{\Delta s},$$

since

$$C = \mathbb{E}[F(X_0)].$$

The following inequalities will be useful, and should be kept in mind to understand what will follow.

Lemma 5.4.20. *i) For almost all ξ, and all x we have :*

$$|\tilde{X}_n^x(T) - X_n^x(T)| \leq C_3 \|v - \tilde{v}\|_\infty.$$

ii) For almost all ξ, and all x we have :

$$|Y_n^x| \leq \|v - \tilde{v}\|_\infty (1 + C_3 V)\Delta s.$$

where we use the notation

$$Y_n^x = X_n^x(T+\Delta s) - X_n^x(T) - (\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)).$$

iii) There exists a constant C_5 depending only on d and V such that for all x, for $p = 1$ and $p = 2$ we have :

$$\|\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)\|_{L^p} \leq C_5\sqrt{\Delta s} \text{ and } \|X_n^x(T+\Delta s) - X_n^x(T)\|_{L^p} \leq C_5\sqrt{\Delta s}.$$

Proof. i) It is a direct application of Proposition 5.4.9.

ii) For almost all ξ, and all x we have, using Proposition 5.4.9 :

$$\begin{aligned}
&|(X_n^x(T+\Delta s) - X_n^x(T)) - (\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))| \\
&= |v(X_n^x(T)) - \tilde{v}(\tilde{X}_n^x)|\Delta s \\
&= |v(X_n^x(T)) - v(\tilde{X}_n^x(T)) + v(\tilde{X}_n^x(T)) - \tilde{v}(\tilde{X}_n^x)|\Delta s \\
&\leq \|Dv\|_\infty C_3 \|v - \tilde{v}\|_\infty \Delta s + \|v - \tilde{v}\|_\infty \Delta s,
\end{aligned}$$

where we have used the previous point.

iii)

$$\begin{aligned}
|X_n^x(T+\Delta s) - X_n^x(T)| &= |v(X_n^x(T))\Delta s + \sqrt{2D\Delta s}N| \\
|\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)| &= |\tilde{v}(\tilde{X}_n^x(T))\Delta s + \sqrt{2D\Delta s}N|,
\end{aligned}$$

where N is a mean-free gaussian whose covariance is equal to identity.
Therefore, using that for h small enough $\|\tilde{v}\| \leq 2V$ by Proposition 5.4.3, we get

$$\begin{aligned}
\|X_n^x(T+\Delta s) - X_n^x(T)\|_{L^1} &\leq V\Delta s + \sqrt{2D\Delta s} \\
\|\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)\|_{L^1} &\leq 2V\Delta s + \sqrt{2D\Delta s} \\
\|X_n^x(T+\Delta s) - X_n^x(T)\|_{L^2} &\leq V\Delta s + \sqrt{2D\Delta s} \\
\|\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)\|_{L^2} &\leq 2V\Delta s + \sqrt{2D\Delta s}
\end{aligned}$$

□

We can now bound F.

$$\begin{aligned} F(x) &= \frac{\mathbb{E}[\varphi(\tilde{X}_n^x(T+\Delta s))] - \mathbb{E}[\varphi(\tilde{X}_n(T))]}{\Delta s} - \frac{\mathbb{E}[\varphi(X_n^x(T+\Delta t))] - \mathbb{E}[\varphi(X_n(T)^x)]}{\Delta s} \\ &= \frac{1}{\Delta s}\mathbb{E}\left[\int_0^1 D\varphi(\tilde{X}_n^x(T)) + u(\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))).(\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))du\right] \\ &- \frac{1}{\Delta s}\mathbb{E}\left[\int_0^1 D\varphi(X_n^x(T)) + u(X_n^x(T+\Delta s) - X_n^x(T))).(X_n^x(T+\Delta s) - X_n(T))du\right] \\ &= f_1(x) + f_2(x), \end{aligned}$$

where f_1 and f_2 are defined by

$$\begin{aligned} f_1(x) &= \frac{1}{\Delta s}\mathbb{E}\left[\int_0^1 D\varphi(\tilde{X}_n^x(T)) + u(\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))).(\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))du\right] \\ &- \frac{1}{\Delta s}\mathbb{E}\left[\int_0^1 D\varphi(X_n^x(T)) + u(X_n^x(T+\Delta s) - X_n^x(T))).(\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n(T))du\right]. \end{aligned}$$

and

$$f_2(x) = \frac{1}{\Delta s}\mathbb{E}\left[\int_0^1 D\varphi[X_n^x(T) + u(X_n^x(T+\Delta s) - X_n^x(T))].Y_n^x du\right]$$

We first treat f_2 : the application of the second point of Lemma 5.4.20 gives that for all x

$$|f_2(x)| \leq \kappa\|v - \tilde{v}\|_\infty(1 + C_3 V).$$

We now treat f_1

$$f_1(x) = \frac{1}{\Delta s}\mathbb{E}\left[\int_{[0,1]^2} D^2\varphi(Z_n^x(u,\lambda)).(\tilde{X}_n^x(T) - X_n^x(T) + uY_n^x, \tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))dud\lambda\right],$$

where we use the notation :

$$\begin{aligned} Z_n^x(u,\lambda) &= \lambda[\tilde{X}_n^x(T) + u(\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)] \\ &+ (1-\lambda)[X_n^x(T) + u(X_n^x(T+\Delta s) - X_n^x(T)]. \end{aligned}$$

We split f_1 into two terms, i.e. $f_1(x) = f_3(x) + f_4(x)$ where we define :

$$f_3(x) = \frac{1}{\Delta s}\mathbb{E}\left[\int_{[0,1]^2} D^2\varphi(Z_n^x(u,\lambda)).(\tilde{X}_n^x(T) - X_n^x(T), \tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))dud\lambda\right],$$

and

$$f_4(x) = \frac{1}{\Delta s}\mathbb{E}\left[\int_{[0,1]^2} D^2\varphi(Z_n^x(u,\lambda)).(uY_n^x, \tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))dud\lambda\right].$$

We first bound f_4 using the second and third points of Lemma 5.4.20.

$$|f_4(x)| \leq \kappa\|v - \tilde{v}\|_\infty(1 + C_3 V)C_5\sqrt{\Delta s} \leq \kappa\|v - \tilde{v}\|_\infty(1 + C_3 V)C_5\sqrt{T}.$$

We now bound f_3, by splitting it into two terms, i.e. $f_3 = f_5 + f_6$, where

$$f_5(x) = \frac{\mathbb{E}\left[\int_{[0,1]^2} D^2\varphi(\lambda\tilde{X}_n^x(T) + (1-\lambda)X_n^x(T)).(\tilde{X}_n^x(T) - X_n^x(T), \tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T))dud\lambda\right]}{\Delta s},$$

and

$$f_6(x) = \frac{\mathbb{E}[\int_{[0,1]^3} D^3\varphi(W_n^x(\lambda,\mu,u)).(Q_n^x(\lambda,u), \tilde{X}_n^x(T) - X_n^x(T), (\tilde{X}_n^x(T+\Delta s - \tilde{X}_n^x(T))))d\mu d\lambda du]}{\Delta s},$$

where

$$Q_n^x(\lambda,u) = \lambda s(\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)) + (1-\lambda)u(X_n^x(T+\Delta s) - X_n^x(T)),$$

and

$$W_n^x(\lambda,\mu) = \lambda\tilde{X}_n^x(T) + (1-\lambda)X_n^x(T) + \mu Q_n^x(\lambda,u).$$

We first bound f_5, by using the independence of $(\lambda\tilde{X}_n^x(T) + (1-\lambda)X_n^x(T), \tilde{X}_n^x(T) - X_n^x(T))$, which is $\mathcal{F}_T$-measurable and $(W(T+\Delta T) - W(T))$ together with the fact that $\mathbb{E}[W(T+\Delta T) - W(T)] = 0$.

$$\begin{aligned} f_5(x) &= \frac{\mathbb{E}\left[\int_{[0,1]^2} D^2\varphi(\lambda\tilde{X}_n^x(T) + (1-\lambda)X_n^x(T)).(\tilde{X}_n^x(T) - X_n^x(T), \tilde{v}(\tilde{X}_n^x(T))\Delta s)dud\lambda\right]}{\Delta s} \\ &+ \frac{\mathbb{E}\left[\int_{[0,1]^2} D^2\varphi(\lambda\tilde{X}_n^x(T) + (1-\lambda)X_n^x(T)).(\tilde{X}_n^x(T) - X_n^x(T), \sqrt{2D}(W(T+\Delta T) - W(T)))dud\lambda\right]}{\Delta s} \\ &= \frac{\mathbb{E}\left[\int_{[0,1]^2} D^2\varphi(\lambda\tilde{X}_n^x(T) + (1-\lambda)X_n^x(T)).(\tilde{X}_n^x(T) - X_n^x(T), \tilde{v}(\tilde{X}_n^x(T))\Delta s)dud\lambda\right]}{\Delta s} \\ &+ 0. \end{aligned}$$

Whence, the first point of Lemma 5.4.20 yields

$$|f_5(x)| \leq \kappa V C_3 \|v - \tilde{v}\|_\infty.$$

Besides this, using the first and third points of Lemma 5.4.20 and Cauchy-Schwarz inequality we get

$$\begin{aligned} |f_6(x)| &\leq \kappa C_3\|v-\tilde{v}\|_\infty \frac{\int_{[0,1]^3} \mathbb{E}\left[|Q_n^x(\lambda,u)||\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)\|\right] d\mu d\lambda du}{\Delta s} \\ &\leq \kappa C_3\|v-\tilde{v}\|_\infty \\ &\quad \frac{(\|\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)\|_{L^2} + \|X_n^x(T+\Delta s) - X_n^x(T)\|_{L^2})\|\tilde{X}_n^x(T+\Delta s) - \tilde{X}_n^x(T)\|_{L^2}}{\Delta s} \\ &\leq 2\kappa C_3\|v-\tilde{v}\|_\infty C_5^2. \end{aligned}$$

Taking the sum of all these bounds yields :

$$\begin{aligned} |F(x)| &\leq \kappa\|v-\tilde{v}\|_\infty(1+C_3V) + \kappa\|v-\tilde{v}\|_\infty(1+C_3V)C_5\sqrt{T} \\ &+ \kappa V C_3\|v-\tilde{v}\|_\infty + 2\kappa C_3\|v-\tilde{v}\|_\infty C_5^2. \end{aligned}$$

Thus,

$$|C| \leq C_6\|v-\tilde{v}\|_\infty.$$

And therefore there exists a constant C_7 such that for almost all ω, we have :

$$|A| \leq C_7(\|v-\tilde{v}\|_\infty + \Delta s + \Delta t). \tag{5.13}$$

We can now bound $E2$.

$$\begin{aligned} E2 &= D\psi(\mathbb{E}[\varphi(X(T))]).\frac{\mathbb{E}[\varphi(X(T+\Delta s))-\varphi(X(T))]}{\Delta s} \\ &\quad -\frac{\psi(\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))])-\psi(\mathbb{E}[\varphi(\tilde{X}_n(T))])}{\Delta s} \\ &= D\psi(\mathbb{E}[\varphi(X(T))]).\frac{\mathbb{E}[\varphi(X(T+\Delta s))-\varphi(X(T))]}{\Delta s} \\ &- D\psi(\mathbb{E}[\varphi(\tilde{X}_n(T))]).\frac{\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))-\varphi(\tilde{X}_n(T))]}{\Delta s} \\ &- \int_0^1 D^2\psi(M_n(u)).\frac{(\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))-\varphi(\tilde{X}_n(T))])^2}{\Delta s}du \end{aligned}$$

where $M_n(u)=\mathbb{E}[\varphi(\tilde{X}_n(T)]+u(\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))-\varphi(\tilde{X}_n(T))]])$. Hence, we have

$$\begin{aligned} E2 &= D\psi(\mathbb{E}[\varphi(X(T))]).A \\ &+ (D\psi(\mathbb{E}[\varphi(X(T))])-D\psi(\mathbb{E}[\varphi(\tilde{X}_n(T))])).\frac{(\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))-\varphi(\tilde{X}_n(T))])}{\Delta s} \\ &- \int_0^1 D^2\psi(M_n(u)).\frac{(\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))-\varphi(\tilde{X}_n(T))])^2}{\Delta s}du. \end{aligned}$$

In order to bound $E2$, we prove the following lemma :

Lemma 5.4.21. *There exists a constant C_8 such that for h small enough, we have for almost all ω*

$$|\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))]-\mathbb{E}[\varphi(\tilde{X}_n(T))]|\leq C_8\Delta s.$$

Proof. We recall that

$$\tilde{X}_n(T+\Delta s)-\tilde{X}_n(T)=\tilde{v}(\tilde{X}_n(T))\Delta s+\sqrt{2D\Delta s}N.$$

$$\begin{aligned} &\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))]-\mathbb{E}[\varphi(\tilde{X}_n(T))] \\ &=\mathbb{E}[D\varphi(\tilde{X}_n(T)).(\tilde{v}(\tilde{X}_n(T))\Delta s+\sqrt{2D\Delta s}N)] \\ &+\mathbb{E}\left[\int_0^1 D^2\varphi(\tilde{X}_n(T)+u(\tilde{X}_n(T+\Delta s)-\tilde{X}_n(T))).(\tilde{v}(\tilde{X}_n(T))\Delta s+\sqrt{2D\Delta s}N)^2du\right] \\ &=\mathbb{E}[D\varphi(\tilde{X}_n(T)).(\tilde{v}(\tilde{X}_n(T))\Delta s)]+0 \\ &+\mathbb{E}\left[\int_0^1 D^2\varphi(\tilde{X}_n(T)+u(\tilde{X}_n(T+\Delta s)-\tilde{X}_n(T))).(\tilde{v}(\tilde{X}_n(T))\Delta s+\sqrt{2D\Delta s}N)^2du\right] \end{aligned}$$

where we have used once again the independence of N and $\tilde{X}_n^x(T)$, whence for h small enough :

$$|\mathbb{E}[\varphi(\tilde{X}_n(T+\Delta s))]-\mathbb{E}[\varphi(\tilde{X}_n(T))]| \leq 2\kappa V\Delta s+\kappa\mathbb{E}[(2V\sqrt{T}+\sqrt{2D}N)^2]\Delta s.$$

□

We can now bound $E2$, using the Lemma 5.4.21, the bound (5.13) of A, the bound (5.12) and the first point of Lemma 5.4.20

$$\begin{aligned} |E2| &\leq \kappa A+\kappa C_8|\mathbb{E}[\varphi(X(T))-\mathbb{E}[\varphi(\tilde{X}_n(T))]|+\kappa C_8^2\Delta s \\ &\leq \kappa C_7(\|v-\tilde{v}\|_\infty+\Delta s+\Delta t)+\kappa C_8(C_2\Delta s+\kappa C_3\|v-\tilde{v}\|_\infty)+\kappa C_8^2\Delta s. \end{aligned}$$

We introduce a new constant C_9 which depends only on v through V and which is therefore independent of ω, and we have then finally

$$|E2|\leq C_9(\Delta s+\Delta t+\|v-\tilde{v}\|_\infty).$$

E3 For $1 \leq i \leq N$ we introduce the notation

$$P_i(\omega) = \frac{\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s))]) - \psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))])}{\Delta s}.$$

The $(P_i)_{1\leq i\leq N}$ are independent and identically distributed L^2 random variables, therefore

$$\begin{aligned} \|E3(\omega)\|_{L^2_\omega} &= \left\| \mathbb{E}[P_i] - \frac{1}{N}\sum_{i=1}^N P_i \right\|_{L^2_\omega} \\ &\leq \frac{\|P_i - \mathbb{E}[P_i]\|_{L^2_\omega}}{\sqrt{N}} \end{aligned}$$

For almost all ω, we have, using Markov property and Lemma 5.4.21

$$\begin{aligned} |P_i(\omega)| &\leq \|D\psi\|_\infty \frac{|\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s))]) - \mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))]|}{\Delta s} \\ &\leq \|D\psi\|_\infty C_8. \end{aligned}$$

Thus

$$\|E3(\omega)\|_{L^2_\omega} \leq \frac{2\kappa C_8}{\sqrt{N}}.$$

E4 We use some preliminary results to bound $E4$.
Let $1 \leq i \leq N$ and ω be fixed, then for $1 \leq j \leq M$, we introduce the random variables

$$Q_{i,j}(\omega,\xi)) = \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T+\Delta s)) - \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T)).$$

We recall that if $(U_j)_{1\leq j\leq M}$ are idependant and identically distributed random variables in L^4, then

$$\left\| \mathbb{E}[U_j] - \frac{1}{M}\sum_{j=1}^M U^j \right\|_{L^4_\xi} \leq \frac{2}{\sqrt{M}}(\|U_j\|_{L^4} + \|U_j\|_{L^2}).$$

The $(\varphi(\tilde{X}_n^{i,j}(\omega,\cdot))_{1\leq j\leq M}$ being independent and identically distributed L^4 random variables, we have

$$\left\| \mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(\omega,\xi,T))] - \frac{1}{M}\sum_{j=1}^M \varphi(\tilde{X}_n^{i,j}(\omega,\xi,T)) \right\|_{L^4_\xi} \leq \frac{4\kappa}{\sqrt{M}}. \tag{5.14}$$

For almost all ω, ξ and all i, j we have

$$\begin{aligned} |Q_{i,j}(\omega,\xi)| &\leq \kappa\left(V\Delta s + \sqrt{2D}(W^{i,j}(T+\Delta s) - W^{i,j}(T))\right) \\ &\leq \kappa\sqrt{\Delta s}\left(V\sqrt{T} + \sqrt{2D}\frac{(W^{i,j}(T+\Delta s) - W^{i,j}(T))}{\sqrt{\Delta s}}\right). \end{aligned}$$

Therefore, for almost all ω and all i we have

$$\|Q_{i,j}(\omega,\xi)\|_{L^2_\xi} \leq C_9\kappa\sqrt{\Delta s},$$

and

$$\|Q_{i,j}(\omega,\xi)\|_{L^4_\xi} \leq C_{10}\kappa\sqrt{\Delta s}.$$

Besides this, the $(Q_{i,j}(\omega,\cdot)_{1\leq j\leq M}$ are idependant and identically distributed random variables in L^4, therefore

$$\left\|\mathbb{E}_\xi[Q_{i,j}(\omega,\xi)]-\frac{1}{M}\sum_{j=1}^{M}Q_{i,j}(\omega,\xi)\right\|_{L^4_\xi}\leq\frac{2}{\sqrt{M}}(C_9+C_{10})\kappa\sqrt{\Delta s},\tag{5.15}$$

and

$$\left\|\mathbb{E}_\xi[Q_{i,j}(\omega,\xi)]-\frac{1}{M}\sum_{j=1}^{M}Q_{i,j}(\omega,\xi)\right\|_{L^2_\xi}\leq\frac{1}{\sqrt{M}}C_9\kappa\sqrt{\Delta s},\tag{5.16}$$

Finally, for almost all ω, ξ and all $1\leq i\leq N$,

$$\begin{aligned}
&\frac{\psi(\mathbb{E}[\varphi(\tilde{X}_n^{i,j}(T+\Delta s))])-\psi(\mathbb{E}[\varphi(\tilde{X}_n^{i,j}(T))])}{\Delta s}\\
&-\frac{\psi(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(T+\Delta s)))-\psi(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(T)))}{\Delta s}\\
&=\int_0^1 D\psi(\mathbb{E}[\varphi(\tilde{X}_n^{i,j}(T))]+u\mathbb{E}[Q_{i,j}]).\frac{\mathbb{E}_\xi[Q_{i,j}]}{\Delta s}du\\
&-\int_0^1 D\psi\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(,T))+\frac{1}{M}\sum_{j=1}^{M}Q_{i,j}\right).\frac{\frac{1}{M}\sum_{j=1}^{M}Q_{i,j}}{\Delta s}du\\
&=\int_0^1 D\psi(\mathbb{E}[\varphi(\tilde{X}_n^{i,j}(T))]+u\mathbb{E}[Q_{i,j}]).\left(\frac{\mathbb{E}_\xi[Q_{i,j}]-\frac{1}{M}\sum_{j=1}^{M}Q_{i,j}}{\Delta s}\right)du\\
&-\int_0^1\left(D\psi(\mathbb{E}[\varphi(\tilde{X}_n^{i,j}(T))+uQ_{i,j}])-D\psi\left(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(T))+uQ_{i,j}\right)\right).\frac{\sum_{j=1}^{M}Q_{i,j}}{M\Delta s}du.
\end{aligned}$$

Thus, thanks to Cauchy-Schwarz inequality and the above preliminary results (5.14), (5.16) and (5.15), we have for almost all ω, ξ and all $1\leq i\leq N$,

$$\begin{aligned}
&\|\frac{\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(T+\Delta s))])-\psi(\mathbb{E}_\xi[\varphi(\tilde{X}_n^{i,j}(T))])}{\Delta s}\\
&-\frac{\psi(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(T+\Delta s)))-\psi(\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(T)))}{\Delta s}\|_{L^2_\xi}\\
&\leq\frac{\kappa}{\Delta s}\left\|\mathbb{E}_\xi[Q_{i,j}]-\frac{1}{M}\sum_{j=1}^{M}Q_{i,j}\right\|_{L^2_\xi}\\
&+\frac{\kappa\left\|Q_{i,j}\right\|_{L^4_\xi}}{\Delta s}\left(\left\|\mathbb{E}[\varphi(\tilde{X}_n^{i,j}(T))]-\frac{1}{M}\sum_{j=1}^{M}\varphi(\tilde{X}_n^{i,j}(T))\right\|_{L^4_\xi}+\left\|\mathbb{E}[Q_{i,j}]-\frac{1}{M}\sum_{j=1}^{M}Q_{i,j})\right\|_{L^4_\xi}\right)\\
&\leq\frac{\kappa}{\Delta s}\frac{C_9\kappa\sqrt{\Delta s}}{\sqrt{M}}+\kappa\left(\frac{4\kappa}{\sqrt{M}}+\frac{2}{\sqrt{M}}(C_9+C_{10})\kappa\sqrt{\Delta s}\right)\frac{C_{10}\kappa\sqrt{\Delta s}}{\Delta s}\\
&\leq C_{11}\kappa^2(1+\kappa)\frac{1}{\sqrt{M\Delta s}}.
\end{aligned}$$

Taking the sum over i of these inequalities and taking the expected value with respect to ω yields

$$\|E4(\omega,\xi)\|_{L^2_{\omega,\xi}}\leq C_{11}\kappa^2(1+\kappa)\frac{1}{\sqrt{M\Delta s}}.$$

The final result on E follows, taking the sum of the bounds of $E1, E2, E3, E4$.

□

Bibliographie

[1] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45(3) :1005–1034, 2007.

[2] I. Babuška, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2) :800–825, 2004.

[3] Ivo Babuška, Börje Andersson, Paul J. Smith, and Klas Levin. Damage analysis of fiber composites. I. Statistical analysis on fiber scale. *Comput. Methods Appl. Mech. Engrg.*, 172(1-4) :27–77, 1999.

[4] A. Barth, C. Schwab, and N. Zollinger. Multi–level Monte Carlo finite element method for elliptic PDE's with stochastic coefficients. SAM Research Reports 2010-18, ETH Zürich, 2010.

[5] A. Bellin, P. Salandin, and A. Rinaldo. Simulation of dispersion in heterogeneous porous formations : Statistics, first-order theories, convergence of computations. *Water Resour. Res.*, 28 :2211–2227, 1992.

[6] Fred Espen Benth and Jon Gjerde. Convergence rates for finite element approximations of stochastic partial differential equations. *Stochastics Stochastics Rep.*, 63(3-4) :313–326, 1998.

[7] Marcel Bieri and Christoph Schwab. Sparse high order FEM for elliptic sPDEs. *Comput. Methods Appl. Mech. Engrg.*, 198(13-14) :1149–1170, 2009.

[8] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[9] Haïm Brezis. *Analyse fonctionnelle*. Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson, Paris, 1983. Théorie et applications. [Theory and applications].

[10] J. Charrier. Strong and weak error estimates for the solutions of elliptic partial differential equations with random coefficients. Technical Report HAL :inria-00490045, Hyper Articles en Ligne, INRIA, 2010. Available at `http ://hal.inria.fr/inria-00490045/en/`.

[11] P. G. Ciarlet. *The finite element method for elliptic problems*, volume 40 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

[12] K.A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 2011. accepted for publication.

[13] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[14] G. Dagan. *Flow and transport in porous formations*. Springer Verlag. 1989.

[15] J.-R. de Dreuzy, A. Beaudoin, and J. Erhel. Asymptotic dispersion in 2d heterogeneous porous media determined by parallel numerical simulation. *Water Resourc. Res.*, 43, 2007.

[16] G. de Marsily, F. Delay, J. Goncalves, P. Renard, V. Teles, and S. Violette. Dealing with spatial heterogeneity. *Hydrogeol. J*, 13 :161–183, 2005.

[17] P. Delhomme. Spatial variability and uncertainty in groundwater flow param- eters, a geostatistical approach. *Water Resourc. Res.*, pages 269–280, 1979.

[18] M. Dentz, H. Kinzelbach, S. Attinger, and W. Kinzelbach. Temporal behavior of a solute cloud in a heterogeneous porous medium : 3. numerical simulations. *Water Resour. Res.*, 7 :1118, 2002.

[19] I. Elishakoff, Y. K. Lin, and L. P. Zhu. *Probabilistic and convex modelling of acoustically excited structures*, volume 39 of *Studies in Applied Mechanics*. Elsevier Science B.V., Amsterdam, 1994.

[20] Isaac Elishakoff and Yongjian Ren. The bird's eye view on finite element method for structures with large stochastic variations. *Comput. Methods Appl. Mech. Engrg.*, 168(1-4) :51–61, 1999.

[21] A. Ern and J.-L. Guermond. *Éléments finis : théorie, applications, mise en œuvre*, volume 36 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2002.

[22] O. G. Ernst, C. E. Powell, D. J. Silvester, and E. Ullmann. Efficient solvers for a linear stochastic galerkin mixed formulation of diffusion problems with random data. *SIAM J. Sci. Comput.*, 2009.

[23] X. Fernique. Regularité des trajectoires des fonctions aléatoires gaussiennes. In *École d'Été de Probabilités de Saint-Flour, IV-1974*, pages 1–96. Lecture Notes in Math., Vol. 480. Springer, Berlin, 1975.

[24] P. Frauenfelder, C. Schwab, and R. A. Todor. Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Engrg.*, 194(2-5) :205–228, 2005.

[25] R. Freeze. A stochastic conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resour. Res.*, 11 :725–741, 1975.

[26] J. Galvis and M. Sarkis. Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity. *SIAM Journal on Numerical Analysis*, 47 :3624–3651, 2009.

[27] Benjamin Ganis, Hector Klie, Mary F. Wheeler, Tim Wildey, Ivan Yotov, and Dongxiao Zhang. Stochastic collocation and mixed finite elements for flow in porous media. *Comput. Methods Appl. Mech. Engrg.*, 197(43-44) :3547–3559, 2008.

[28] L. W. Gelhar. *Stochastic Subsurface Hydrology*. Engelwood Cliffs. 1993.

[29] R. Ghanem. Ingredients for a general purpose stochastic finite elements implementation. *Comput. Methods Appl. Mech. Engrg.*, 168(1-4) :19–34, 1999.

[30] R. Ghanem and J. Red-Horse. Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach. *Phys. D*, 133(1-4) :137–144, 1999. Predictability : quantifying uncertainty in models of complex phenomena (Los Alamos, NM, 1998).

[31] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements : a spectral approach*. Springer-Verlag, New York, 1991.

[32] R. G. Ghanem and P. D. Spanos. Spectral techniques for stochastic finite elements. *Arch. Comput. Methods Engrg.*, 4(1) :63–100, 1997.

[33] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[34] M. B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. volume 256 of *Monte Carlo and Quasi-Monte Carlo methods 2006*, pages 343–358. Springer, 2007.

[35] M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 256 :981–986, 2008.

[36] C. J. Gittelson. Stochastic galerkin discretization of the log-normal isotropic diffusion problem. *Math. Models Methods Appl. Sci.*, 20(2) :237–263, 2010.

[37] C. Graham, Th. G. Kurtz, S. Méléard, Ph. E. Protter, M. Pulvirenti, and D. Talay. *Probabilistic models for nonlinear partial differential equations*, volume 1627 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1996. Lectures given at the 1st Session and Summer School held in Montecatini Terme, May 22–30, 1995, Edited by Talay and L. Tubaro, Fondazione C.I.M.E.. [C.I.M.E. Foundation].

[38] I.G. Graham, F.Y. Kuo, D. Nuyens, R. Scheichl, and I.H. Sloan. Quasi-monte carlo methods for elliptic pdes with random coefficients and applications. *Journal of Computational Physics*, 230(10) :3668 – 3694, 2011.

[39] W. Hackbusch. *Elliptic differential equations : Theorey and numerical treatment*, volume 18 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2010.

[40] S. Heinrich. Multilevel Monte Carlo methods. volume 2179 of *Lecture notes in Computer Science*, pages 3624–3651. Springer, Philadelphia, PA, 2001.

[41] Einar Hille. Contributions to the theory of Hermitian series. II. The representation problem. *Trans. Amer. Math. Soc.*, 47 :80–94, 1940.

[42] R.J. Hoeksema and P.K. Kitanidis. Analysis of the spatial structure of properties of selected aquifers. *Water Resour. Res.*, 21 :536–572, 1985.

[43] R. Scheichl J. Charrier and A. Teckentrup. Finite element error analysis of elliptic pdes with random coefficients and its application to multilevel monte carlo methods. *submitted*, 2011.

[44] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[45] Michael Kleiber and Tran Duong Hien. *The stochastic finite element method*. John Wiley & Sons Ltd., Chichester, 1992. Basic perturbation technique and computer implementation, With a separately available computer disk.

[46] O. P. Le Maître, O. M. Knio, H. N. Najm, and R. G. Ghanem. Uncertainty propagation using Wiener-Haar expansions. *J. Comput. Phys.*, 197(1) :28–57, 2004.

[47] O. P. Le Maître, H. N. Najm, P. P. Pébay, R. G. Ghanem, and O. M. Knio. Multi-resolution-analysis scheme for uncertainty quantification in chemical systems. *SIAM J. Sci. Comput.*, 29(2) :864–889 (electronic), 2007.

[48] Wing Kam Liu, Ted Belytschko, and A. Mani. Random field finite elements. *Internat. J. Numer. Methods Engrg.*, 23(10) :1831–1845, 1986.

[49] M. Loève. *Probability theory. I.* Springer-Verlag, New York, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45.

[50] M. Loève. *Probability theory. II.* Springer-Verlag, New York, fourth edition, 1978. Graduate Texts in Mathematics, Vol. 46.

[51] Zhiming Lu and Dongxiao Zhang. A comparative study on uncertainty quantification for flow in randomly heterogeneous media using Monte Carlo simulations and conventional and KL-based moment-equation approaches. *SIAM J. Sci. Comput.*, 26(2) :558–577 (electronic), 2004.

[52] A. Lunardi. *Analytic semigroups and optimal regularity in parabolic problems.* Progress in Nonlinear Differential Equations and their Applications, 16. Birkhäuser Verlag, Basel, 1995.

[53] Hermann G. Matthies. Stochastic finite elements : computational approaches to stochastic partial differential equations. *ZAMM Z. Angew. Math. Mech.*, 88(11) :849–873, 2008.

[54] Hermann G. Matthies and Christian Bucher. Finite elements for stochastic media problems. *Comput. Methods Appl. Mech. Engrg.*, 168(1-4) :3–17, 1999.

[55] R. Mikulevičius and E. Platen. Rate of convergence of the Euler approximation for diffusion processes. *Math. Nachr.*, 151 :233–239, 1991.

[56] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics.* Scientific Computation. Springer-Verlag, Berlin, 2004.

[57] S. P. Neumann. Eulerian-lagrangian theory of transport in space-time nonstationary velocity fields : Exact nonlocal formalism by conditional moments and weak approximations. *Water Resour. Res.*, 29 :633–645, 1993.

[58] F. Nobile, R. Tempone, and C. G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5) :2411–2442, 2008.

[59] F. Nobile, R. Tempone, and C. G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5) :2309–2345, 2008.

[60] F. Riesz and B. Sz.-Nagy. *Functional analysis.* Dover Books on Advanced Mathematics. Dover Publications Inc., New York, 1990. Translated from the second French edition by Leo F. Boron, Reprint of the 1955 original.

[61] J. B. Roberts and P.-T. D. Spanos. *Random vibration and statistical linearization.* John Wiley & Sons Ltd., Chichester, 1990.

[62] Y. Rubin. Stochastic modeling of macrodispersion in heterogeneous porous media. *Water Resour. Res.*, 26 :133–141, 1990.

[63] P. Salandin and V. Fiorotto. Dispersion tensor evaluation in heterogeneous media for finite peclet values. *Water Resour. Res.*, 36 :1449–1455, 2000.

[64] Christoph Schwab and Radu Alexandru Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, 217(1) :100–122, 2006.

[65] H. Schwarze, U. Jaekel, and H. Vereecken. Estimation of macrodispersion by different approximation methods for flow and transport in randomly heterogeneous media. *Transp. Porous Media*, 43 :265–287, 2001.

[66] A. Stuart and M. Dashti. Uncertainty quantification and weak approximation of an elliptic inverse problem. Technical Report arXiv :1102.0143, `arXiv.org`, 2010. Submitted to SIAM Journal on Numerical Analysis.

[67] N. Suciu, C. Vamos, and K. Sabelfeld. Ergodic simulations for diffusion in random velocity fields. *Monte Carlo and Quasi-Monte Carlo methods*, pages 659–668, 2007.

[68] J. V. Uspensky. On the convergence of quadrature formulas related to an infinite interval. *Trans. Amer. Math. Soc.*, 30(3) :542–559, 1928.

[69] Tobias von Petersdorff and Christoph Schwab. Sparse finite element methods for operator equations with stochastic data. *Appl. Math.*, 51(2) :145–180, 2006.

[70] X. Wan and G. E. Karniadakis. Solving elliptic problems with non-Gaussian spatially-dependent random coefficients. *Comput. Methods Appl. Mech. Engrg.*, 198(21-26) :1985–1995, 2009.

[71] D. Xiu and G. E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2) :619–644 (electronic), 2002.

[72] D. Xiu and G. E. Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.*, 187(1) :137–167, 2003.

[73] Dongbin Xiu. Efficient collocational approach for parametric uncertainty analysis. *Commun. Comput. Phys.*, 2(2) :293–309, 2007.

[74] Dongbin Xiu. Fast numerical methods for stochastic computations : a review. *Commun. Comput. Phys.*, 5(2-4) :242–272, 2009.

[75] Dongbin Xiu and Jan S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, 27(3) :1118–1139 (electronic), 2005.

[76] Dongxiao Zhang and Zhiming Lu. An efficient, high-order perturbation approach for flow in random porous media via karhunen-loève and polynomial expansions. *J. Comput. Phys.*, 194(2) :773–794, 2004.

[77] S. Zhang. Analysis of finite element domain embedding methods for curved domains using uniform grids. *SIAM Journal on Numerical Analysis*, 46 :2843–2866, 2008.

Résumé

Ce travail présente quelques résultats concernant des méthodes numériques déterministes et probabilistes pour des équations aux dérivées partielles à coefficients aléatoires, avec des applications à l'hydrogéologie.

On s'intéresse tout d'abord à l'équation d'écoulement dans un lieu poreux en régime stationnaire avec un coefficient de perméabilité lognormal homogène, incluant le cas d'une fonction de covariance peu régulière. On établit des estimations aux sens fort et faible de l'erreur commise sur la solution en tronquant le développement de Karhunen-Loève du coefficient. Puis on établit des estimations d'erreurs éléments finis dont on déduit une extension de l'estimation d'erreur existante pour la méthode de collocation stochastique, ainsi qu'une estimation d'erreur pour une méthode de Monte-Carlo multi-niveaux.

On s'intéresse enfin au couplage de l'équation d'écoulement considérée précédemment avec une équation d'advection-diffusion, dans le cas d'incertitudes importantes et d'une faible longueur de corrélation. On propose l'analyse numérique d'une méthode numérique pour calculer la vitesse moyenne à laquelle la zone contaminée par un polluant s'étend. Il s'agit d'une méthode de Monte-Carlo combinant une méthode d'élements finis pour l'équation d'écoulement et un schéma d'Euler pour l'équation différentielle stochastique associée à l'équation d'advection-diffusion, vue comme une équation de Fokker-Planck.

Mots clés : quantification des incertitudes, coefficient lognormal, développement de Karhunen-Loève, méthode d'éléments finis, méthode de collocation stochastique, méthode de Monte-Carlo multi-niveaux, méthode de Monte-Carlo, schéma d'Euler pour des équations différentielles stochastiques.

Abstract

This work presents some results about probabilistic and deterministic numerical methods for partial differential equations with random coefficients, with applications to hydrogeology.

We first consider the steady flow equation in porous media with a homogeneous lognormal permeability coefficient, including the case of a low regularity covariance function. We establish error estimates, both in strong and weak senses, of the error in the solution resulting from the truncature of the Karhunen-Loève expansion of the coefficient. Then we establish finite element error estimates, from which we deduce an extension of the existing error estimate for the stochastic collocation method along with an error estimate for a multilevel Monte-Carlo method.

We finally consider the coupling of the previous flow equation with an advection-diffusion equation, in the case when the uncertainty is important and the correlation length is small. We propose the numerical analysis of a numerical method, which aims at computing the mean velocity of the expansion of a pollutant. The method consists in a Monte-Carlo method, combining a finite element ethod for the flow equation and an Euler scheme for the stochastic differential equation associated to the advection-diffusion equation, seen as a Fokker-Planck equation.

Keywords: uncertainty quantification, lognormal coefficient, Karhunen-Loève expansion, finite element method, stochastic collocation method, multilevel Monte-Carlo method, Monte-Carlo method, Euler scheme for stochastic differential equations.

N° d'ordre :

École normale supérieure de Cachan - Antenne de Bretagne

Campus de Ker Lann - Avenue Robert Schuman - 35170 BRUZ

Tél : +33(0)2 99 05 93 00 - Fax : +33(0)2 99 05 93 29